GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 02:47:58 ; Search time 6394 Seconds
(without alignments)
17232.284 Million cell updates/sec

Title: US-09-715-876-7
Perfect score: 3786
Sequence: 1 atgcttcaacaatttacatt.....tgtccttgttcatttagtga 3786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3786	100.0	3786	8	YSAALS1	L25902 Candida alb
2	1834.4	48.5	3360	8	CAU87956	U87956 Candida alb
3	1708.6	45.1	4383	8	AF051313	AF051313 Candida a
4	1526.4	40.3	4569	8	AF272027	AF272027 Candida a
5	1392.8	36.8	4723	8	AF025429	AF025429 Candida a
6	1188.8	31.4	3813	8	AF068866	AF068866 Candida a
7	1069.4	28.2	1071	6	AR044076	AR044076 Sequence
8	1069.4	28.2	1071	6	I65632	I65632 Sequence 1
9	1016.6	26.9	1404	8	CANALS2S1	AF024582 Candida a
10	1006	26.6	1404	8	CA1ALS1	AF024580 Candida a
11	905.8	23.9	4332	8	AF075293	AF075293 Candida a
12	874.4	23.1	6897	8	AF201684	AF201684 Candida a
13	842.8	22.3	1407	8	CANALS4S1	AF024584 Candida a
14	836.4	22.1	1407	8	CANALS42S1	AF024586 Candida a
15	805.4	21.3	1404	8	AF229989S1	AF229989 Candida a
16	731.8	19.3	1007	8	AF202529	AF202529 Candida d
17	507	13.4	995	8	AF202530	AF202530 Candida d
18	463.6	12.2	1013	8	AF201685	AF201685 Candida d
19	394	10.4	1299	8	AF075294	AF075294 Candida a
20	362.2	9.6	1058	8	AF201686	AF201686 Candida t
21	348.2	9.2	425	6	AR044080	AR044080 Sequence
22	348.2	9.2	425	6	I65636	I65636 Sequence 5
23	331.2	8.7	680	8	AF189016	AF189016 Candida a
24	202.4	5.3	1400	8	AB002099	AB002099 Candida t
25	198.6	5.2	13370	8	SPAPB2C8	AL590602 S.pombe c
c 26	170.8	4.5	172307	2	AC044842	AC044842 Homo sapi
27	141.6	3.7	377	8	AF211866	AF211866 Candida t
c 28	135.8	3.6	132449	9	AL365272	AL365272 Human DNA
c 29	134.4	3.5	20325	8	SPCPB16A4	AL591677 S.pombe c
30	132	3.5	380	8	AF211865	AF211865 Candida t
c 31	130.4	3.4	158615	2	AC117835	AC117835 Rattus no
c 32	126	3.3	185994	2	AC002042	AC002042 Homo sapi
c 33	125.8	3.3	175748	2	AC120669	AC120669 Rattus no
c 34	125.8	3.3	180903	2	AC125859	AC125859 Rattus no
35	121.6	3.2	21251	1	U23947	U23947 Mycoplasma
36	116	3.1	53352	5	AL592077	AL592077 Zebrafish
37	112.2	3.0	157	8	AF035757	AF035757 Candida a
38	111.8	3.0	180668	2	AC020857	AC020857 Mus muscu
c 39	111.6	2.9	666	8	AF413050S2	AF413051 Zea mays
40	111.2	2.9	22398	5	FRU271723	AJ271723 Fugu rubr
c 41	110.2	2.9	141017	2	AC116962	AC116962 Dictyoste
c 42	109.8	2.9	35412	8	SPCC188	AL049662 S.pombe c
c 43	109.8	2.9	38141	8	SPAC8A4	Z66569 S.pombe chr
44	109.6	2.9	127354	2	AC117014	AC117014 Rattus no
45	109.6	2.9	155019	2	AC117361	AC117361 Rattus no

ALIGNMENTS

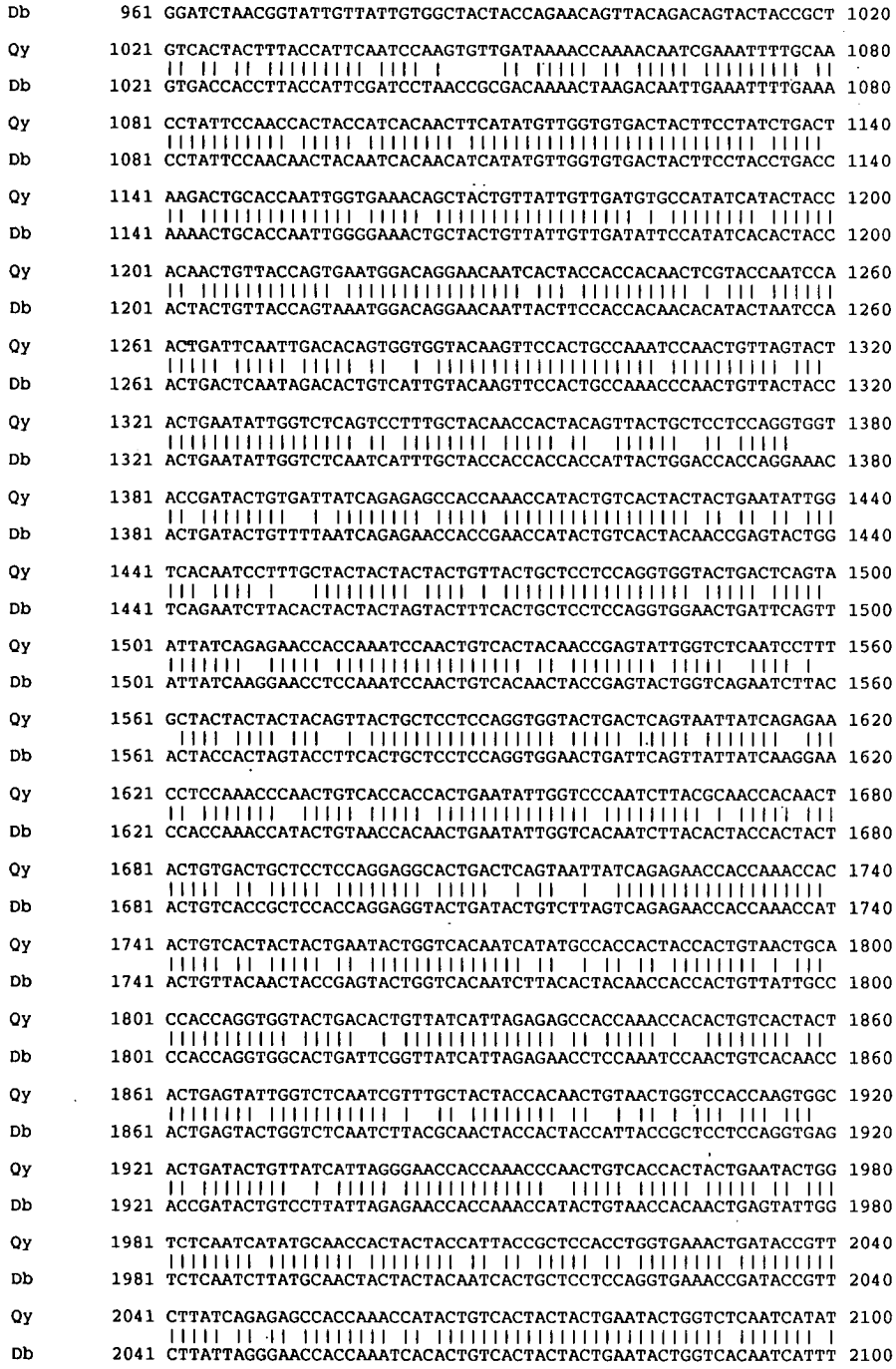
RESULT 1
YSAALS1
LOCUS YSAALS1 3786 bp DNA linear PLN 03-MAY-2000
DEFINITION Candida albicans agglutinin-like sequence (ALS1) gene, complete cds.
ACCESSION L25902
VERSION L25902.1 GI:704426
KEYWORDS .
SOURCE Candida albicans.
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 3786)
AUTHORS Hoyer,L.L., Scherer,S., Shatzman,A.R. and Livi,G.P.
TITLE Candida albicans ALS1: domains related to a Saccharomyces

```
          cerevisiae sexual agglutinin separated by a repeating motif
  JOURNAL   Mol. Microbiol. 15 (1), 39-54 (1995)
  MEDLINE   95272392
   PUBMED   7752895
FEATURES             Location/Qualifiers
     source          1. .3786
                     /organism="Candida albicans"
                     /strain="B792"
                     /db_xref="taxon:5476"
     gene            1. .3783
                     /gene="ALS1"
     CDS             1. .3783
                     /gene="ALS1"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAC41649.2"
                     /db_xref="GI:7687905"
                     /translation="MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPG
                     YPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGE
                     EFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTF
                     NDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSS
                     SNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFI
                     DAYISATDVNQYTLAYTNDYTCAGSRSQSKPFTLRWTGYKNSDAGSNGIVIVATTRTV
                     TDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDVPYHTTTTVTSEWTGTITTTTTTRTNPTDSIDTVVVQVPSPNPTVSTTEYWSQSFAT
                     TTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNP
                     TVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPP
                     GGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW
                     SQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLI
                     REPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTT
                     TVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSK
                     ISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTES
                     ETTTESICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISAT
                     SSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGD
                     NQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSIT
                     EDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENS
                     GLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSESVATTQ
                     ATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTN
                     SNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI"
     repeat_region   1299. .2370
                     /rpt_unit=1299. .1407
BASE COUNT     1170 a    899 c    592 g   1125 t
ORIGIN

  Query Match           100.0%;  Score 3786;  DB 8;  Length 3786;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 3786;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCTTCAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCTTCAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATC 60

Qy         61 ACTGGTGTTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACTGGTGTTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTC 120

Qy        121 AAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGT 180

Qy        181 GCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCA 240

Qy        241 CAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 300

Qy        301 GGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 360

Qy        361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 420

Qy        421 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTT 480

Qy        481 AATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCA 540

Qy        541 AGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTT 600

Qy        601 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC 660

Qy        661 GGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGAT 720

Qy        721 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 780

Qy        781 ATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 ATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATT 840

Qy        841 TCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 TCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGC 900

Qy        901 AGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCC 960

Qy        961 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1020

Qy       1021 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1080

Qy       1081 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1140

Qy       1141 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1200

Qy       1201 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCA 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCA 1260

Qy       1261 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320

Qy       1321 ACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGT 1380

Qy       1381 ACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATATTGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 ACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATATTGG 1440

Qy       1441 TCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTA 1500
```

```
Db      1441 TCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTA 1500
Qy      1501 ATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTT 1560
Db      1501 ATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTT 1560
Qy      1561 GCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAA 1620
Db      1561 GCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAA 1620
Qy      1621 CCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACT 1680
Db      1621 CCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACT 1680
Qy      1681 ACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAACCACCAAACCAC 1740
Db      1681 ACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAACCACCAAACCAC 1740
Qy      1741 ACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACCACTACCACTGTAACTGCA 1800
Db      1741 ACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACCACTACCACTGTAACTGCA 1800
Qy      1801 CCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCACCAAACCACACTGTCACTACT 1860
Db      1801 CCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCACCAAACCACACTGTCACTACT 1860
Qy      1861 ACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGC 1920
Db      1861 ACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGC 1920
Qy      1921 ACTGATACTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACTACTGAATACTGG 1980
Db      1921 ACTGATACTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACTACTGAATACTGG 1980
Qy      1981 TCTCAATCATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGTT 2040
Db      1981 TCTCAATCATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGTT 2040
Qy      2041 CTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATAT 2100
Db      2041 CTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATAT 2100
Qy      2101 GCTACAACCACCACTGTTACTGCACCACCTGGTGAAACCGATACCGTTCTTATCAGAGAG 2160
Db      2101 GCTACAACCACCACTGTTACTGCACCACCTGGTGAAACCGATACCGTTCTTATCAGAGAG 2160
Qy      2161 CCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACC 2220
Db      2161 CCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACC 2220
Qy      2221 ACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGAGCCACCAAATCCA 2280
Db      2221 ACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGAGCCACCAAATCCA 2280
Qy      2281 ACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACAGTTACTGCT 2340
Db      2281 ACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACAGTTACTGCT 2340
Qy      2341 CCTCCAGGTGGTACTGACACTGTGATTATCTATGAAAGCATGTCAAGTTCAAAGATTTCT 2400
Db      2341 CCTCCAGGTGGTACTGACACTGTGATTATCTATGAAAGCATGTCAAGTTCAAAGATTTCT 2400
Qy      2401 ACATCCTCCAATGATATAACCAGTATCATTCCATCATTTTCCCGTCCTCATTATGTCAAC 2460
Db      2401 ACATCCTCCAATGATATAACCAGTATCATTCCATCATTTTCCCGTCCTCATTATGTCAAC 2460
Qy      2461 AGCACAACCTCCGATTTGTCAACATTTGAATCTTCATCCATGAATACTCCTACTTCTATC 2520
Db      2461 AGCACAACCTCCGATTTGTCAACATTTGAATCTTCATCCATGAATACTCCTACTTCTATC 2520
Qy      2521 AGTAGTGATGGTATGTTGTTGTCTTCTACAACTTTGGTTACTGAATCAGAAACAACTACA 2580
```

```
Db      2521 AGTAGTGATGGTATGTTGTTGTCTTCTACAACTTTGGTTACTGAATCAGAAACAACTACA 2580
Qy      2581 GAACTGATTTGCAGTGATGGTAAAGAGTGTTCTAGATTGTCCAGTTCTTCTGGTATTGTC 2640
Db      2581 GAACTGATTTGCAGTGATGGTAAAGAGTGTTCTAGATTGTCCAGTTCTTCTGGTATTGTC 2640
Qy      2641 ACAAATCCAGATAGCAATGAATCCTCAATCGTAACTAGTACTGTTCCTACTGCAAGTACA 2700
Db      2641 ACAAATCCAGATAGCAATGAATCCTCAATCGTAACTAGTACTGTTCCTACTGCAAGTACA 2700
Qy      2701 ATGTCTGATTCACTTTCTTCAACTGATGGTATTAGTGCTACATCTTCTGATAATGTTTCA 2760
Db      2701 ATGTCTGATTCACTTTCTTCAACTGATGGTATTAGTGCTACATCTTCTGATAATGTTTCA 2760
Qy      2761 AAATCAGGAGTATCAGTTACAACCGAAACTTCTGTTACAACTATTCAAACTACTCCAAAC 2820
Db      2761 AAATCAGGAGTATCAGTTACAACCGAAACTTCTGTTACAACTATTCAAACTACTCCAAAC 2820
Qy      2821 CCATTATCATCTTCAGTGACATCATTGACTCAGTTGTCTTCAATTCCAAGTGTTTCAGAA 2880
Db      2821 CCATTATCATCTTCAGTGACATCATTGACTCAGTTGTCTTCAATTCCAAGTGTTTCAGAA 2880
Qy      2881 AGTGAAAGTAAAGTTACATTTACAAGCAATGGAGACAACCAAAGTGGTACTCATGATTCA 2940
Db      2881 AGTGAAAGTAAAGTTACATTTACAAGCAATGGAGACAACCAAAGTGGTACTCATGATTCA 2940
Qy      2941 CAATCTACTTCCACTGAAATTGAAATTGTAACAACCAGTTCTACTAAAGTTTTACCACCT 3000
Db      2941 CAATCTACTTCCACTGAAATTGAAATTGTAACAACCAGTTCTACTAAAGTTTTACCACCT 3000
Qy      3001 GTCGTTTCTTCTAATACTGATTTGACTAGTGAACCAACAAATACCAGAGAACAACCAACT 3060
Db      3001 GTCGTTTCTTCTAATACTGATTTGACTAGTGAACCAACAAATACCAGAGAACAACCAACT 3060
Qy      3061 ACATTATCAACTACTTCAAACTCCATCACTGAAGATATCACCACATCTCAACCTACAGGT 3120
Db      3061 ACATTATCAACTACTTCAAACTCCATCACTGAAGATATCACCACATCTCAACCTACAGGT 3120
Qy      3121 GATAATGGAGACAATACTTCATCAACCAATCCAGTTCCAACTGTGGCAACAAGTACTTTA 3180
Db      3121 GATAATGGAGACAATACTTCATCAACCAATCCAGTTCCAACTGTGGCAACAAGTACTTTA 3180
Qy      3181 GCATCTGCAAGTGAAGAAGACAACAAAAGCGGTTCTCATGAATCAGCATCCACAAGTTTG 3240
Db      3181 GCATCTGCAAGTGAAGAAGACAACAAAAGCGGTTCTCATGAATCAGCATCCACAAGTTTG 3240
Qy      3241 AAACCAAGTATGGGTGAAAATTCTGGATTAACTACTTCTACTGAAATTGAAGCTACAACA 3300
Db      3241 AAACCAAGTATGGGTGAAAATTCTGGATTAACTACTTCTACTGAAATTGAAGCTACAACA 3300
Qy      3301 ACCAGTCCTACAGAAGCTCCATCACCTGCTGTTTCTTCTGGTACTGATGTAACTACTGAA 3360
Db      3301 ACCAGTCCTACAGAAGCTCCATCACCTGCTGTTTCTTCTGGTACTGATGTAACTACTGAA 3360
Qy      3361 CCAACTGATACTAGAGAACAACCTACTACATTATCAACTACTTCAAAAACAAACAGTGAA 3420
Db      3361 CCAACTGATACTAGAGAACAACCTACTACATTATCAACTACTTCAAAAACAAACAGTGAA 3420
Qy      3421 CTGGTTGCTACTACACAAGCTACTAATGAAAATGGTGGTAAATCTCCATCAACTGATTTA 3480
Db      3421 CTGGTTGCTACTACACAAGCTACTAATGAAAATGGTGGTAAATCTCCATCAACTGATTTA 3480
Qy      3481 ACATCAAGCTTGACAACAGGCACCTCAGCATCTACAAGTGCTAATAGCGAACTTGTTACT 3540
Db      3481 ACATCAAGCTTGACAACAGGCACCTCAGCATCTACAAGTGCTAATAGCGAACTTGTTACT 3540
Qy      3541 AGTGGATCTGTTACTGGTGGAGCTGTTGCCAGTGCTTCAAATGATCAATCACATTCTACT 3600
Db      3541 AGTGGATCTGTTACTGGTGGAGCTGTTGCCAGTGCTTCAAATGATCAATCACATTCTACT 3600
Qy      3601 TCTGTTACCAACAGCAACAGCATTGTATCTAATACCCCACAAACTACATTGAGTCAACAA 3660
Db      3601 TCTGTTACCAACAGCAACAGCATTGTATCTAATACCCCACAAACTACATTGAGTCAACAA 3660
```

```
Qy       3661 GTTACCTCATCCTCACCTTCAACCAACACATTCATTGCTTCTACATACGATGGCTCTGGT 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 GTTACCTCATCCTCACCTTCAACCAACACATTCATTGCTTCTACATACGATGGCTCTGGT 3720

Qy       3721 TCTATTATCCAACATTCTACTTGGTTGTACGGTTTGATCACATTATTGTCCTTGTTCATT 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 TCTATTATCCAACATTCTACTTGGTTGTACGGTTTGATCACATTATTGTCCTTGTTCATT 3780

Qy       3781 TAGTGA 3786
              ||||||
Db       3781 TAGTGA 3786


RESULT 2
CAU87956
LOCUS       CAU87956               3360 bp    DNA     linear   PLN 02-JUL-1998
DEFINITION  Candida albicans agglutinin-like protein (ALS3) gene, complete cds.
ACCESSION   U87956
VERSION     U87956.1  GI:3273414
KEYWORDS    .
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 3360)
  AUTHORS   Hoyer,L.L., Payne,T.L., Bell,M., Myers,A.M. and Scherer,S.
  TITLE     Candida albicans ALS3 and insights into the nature of the ALS gene
            family
  JOURNAL   Curr. Genet. 33 (6), 451-459 (1998)
  MEDLINE   98309840
   PUBMED   9644209
REFERENCE   2  (bases 1 to 3360)
  AUTHORS   Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-1997) Veterinary Pathobiology, University of
            Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES             Location/Qualifiers
     source          1. .3360
                     /organism="Candida albicans"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="R"
                     /map="S+U"
     gene            1. .3360
                     /gene="ALS3"
     CDS             1. .3360
                     /gene="ALS3"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAC39486.1"
                     /db_xref="GI:3273415"
                     /translation="MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYNYKGPG
                     TPTWNAVLGWSLDGTSASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGE
                     EFMTFSTLTCTVSNTLTPSIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTF
                     NDGGKKISINVDFERSNVDPKGYLTDSRVIPSLNKVSTLFVAPQCANGYTSGTMGFAN
                     TYGDVQIDCSNIHVGITKGLNDWNYPVSSESFSYTKTCSSNGIFITYKNVPAGYRPFV
                     DAYISATDVNSYTLSYANEYTCAGGYWQRAPFTLRWTGYRNSDAGSNGIVIVATTRTV
                     TDSTTAVTTLPFDPNRDKTKTIEILKPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDIPYHTTTTVTSKWTGTITSTTTHTNPTDSIDTVIVQVPSPNPTVTTTEYWSQSFAT
                     TTTITGPPGNTDTVLIREPPNHTVTTTEYWSESYTTTSTFTAPPGGTDSVIIKEPPNP
                     TVTTTEYWSESYTTTSTFTAPPGGTDSVIIKEPPNHTVTTTEYWSQSYTTTTTVTAPP
                     GGTDTVLVREPPNHTVTTTEYWSQSYTTTTTVIAPPGGTDSVIIREPPNPTVTTTEYW
                     SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTITAPPGETDTVLI
                     REPPNHTVTTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSYATTT
                     TITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTIIAPPGETDTVLIREPPNPTV
                     TTTEYWSQSYTTATTVTAPPGGTDTVIIYDTMSSSEISSFSRPHYTNHTTLWSTTWVI
                     ETKTITETSCEGDKGCSWVSVSTRIVTIPNNIETPMVTNTVDSTTTESTSQSPSGIFS
                     ESGVSVETESSTVTTAQTNPSVPTTESEVVFTTKGNNENGPYESPSTNVKSSMDENSE
                     FTTSTAASTSTDIENETIATTGSVEASSPIISSSADETTTVTTTAESTSVIEQPTNNN
                     GGGKAPSATSSPSTTTTANNDSVITGTTSTNQSQSQSQYNSDTQQTTLSQQMTSSLVS
                     LHMLTTFDGSGSVIQHSTWLCGLITLLSLFI"

BASE COUNT     1029 a    804 c    520 g   1007 t
ORIGIN

  Query Match            48.5%;  Score 1834.4;  DB 8;  Length 3360;
  Best Local Similarity  73.9%;  Pred. No. 0;
  Matches 2393;  Conservative    0;  Mismatches  830;  Indels   15;  Gaps    5;

Qy          1 ATGCTTCAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATC 60
              ||||| ||||||| |||||||||| || |||||||||  |||| | ||||||||||||
Db          1 ATGCTACAACAATATACATTGTTACTCATATATTTGTCGGTTGCGACTGCAAAGACAATC 60

Qy         61 ACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTC 120
              |||||||||||  | |||||||||||||| |||||||| |||||||||  ||||  ||
Db         61 ACTGGTGTTTTCAACAGTTTTAATTCATTGACTTGGTCTAATGCTGCTACTTATAATTAT 120

Qy        121 AAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGT 180
              ||.|| ||||||  |||||||||||||||||||||||||||||| ||||||||||| |||
Db        121 AAGGGACCAGGAACCCCAACTTGGAATGCTGTTTTGGGTTGGTCTTTAGATGGTACTAGT 180

Qy        181 GCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCA 240
              || | ||| || |||||||||||||||||||||||||||||||||||| |||||||||
Db        181 GCAAGTCCGGGAGATACATTCACATTGAATATGCCATGTGTGTTTAAATTTACTACTTCT 240

Qy        241 CAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 300
              |||||||||||||||| |||||  |||||||||||||||||| |||||||| |  |
Db        241 CAAACATCTGTTGATTTGACTGCTCATGGTGTTAAATATGCTACATGTCAATTTCAGGCA 300

Qy        301 GGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 360
              |||||||||| |  || |||||||||||||||||||||||| | |  ||||||  ||
Db        301 GGTGAAGAATTTATGACCTTTTCTACATTAACATGTACTGTGAGCAATACTTTGACTCCA 360

Qy        361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 420
              || |||||||| || ||||| || || |||||| |||||||||||| ||||||| |||
Db        361 TCTATTAAGGCTTTGGGTACTGTCACCTTACCACTTGCATTCAATGTAGGTGGAACTGGT 420

Qy        421 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTT 480
              || ||   |||||||||||||||||||||||||||||||||| || ||||| ||||||
Db        421 TCTTCTGTTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACTAACACAGTTACATTT 480

Qy        481 AATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCA 540
              |||||||||| ||  | ||||| ||| ||||||| ||||||| |||||  || |||||
Db        481 AATGATGGTGGCAAGAAAATCTCTATTAATGTTGATTTTGAAAGGTCAAATGTCGATCCA 540

Qy        541 AGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTT 600
              |  |  || ||   || |||||||||||| |||||||||| || ||  ||||||||||
Db        541 AAAGGGTACTTAACTGATTCCAGAGTTATACCAAGTCTCAACAAAGTGTCAACTCTTTTT 600

Qy        601 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC 660
              || |||||||||||| |||||||||||||||||||||||||| ||| | |  | | |
Db        601 GTTGCACCACAATGTGCAAATGGTTACACATCTGGTACAATGGGATTCGCTAACACTTAT 660

Qy        661 GGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGAT 720
              ||||| |||   ||||| || |||||||||||.|||||| |||||||||| ||||||
Db        661 GGTGATGTTCAAATTGACTGTTCAAATATTCATGTTGGTATTACAAAAGGATTGAATGAT 720

Qy        721 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 780
              |||||||||||||||||||||||||||||||||||||||||| |||||||| |||||||||
Db        721 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACCAAAACTTGTTCATCTAATGGT 780

Qy        781 ATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATT 840
              ||    || | |||| ||||||| ||||| |||||||||||||| |||| |||||||||
Db        781 ATCTTTATCACATATAAAAATGTTCCTGCCGGTTATCGTCCATTTGTTGACGCTTATATT 840

Qy        841 TCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGC 900
              ||||||||||||||||||    || || ||  | ||| | ||||| ||||||||||||
Db        841 TCTGCTACAGATGTTAATTCGTACACCTTGTCGTATGCTAATGAATATACTTGTGCTGGT 900

Qy        901 AGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCC 960
              ||  |  |||| ||  |||||||| ||||||||||||||||||  ||||||||||
Db        901 GGTTATTGGCAACGTGCACCTTTCACATTAAGATGGACTGGATACAGAAATAGTGATGCT 960

Qy        961 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1020
              |||||||||||||||| ||||| ||||| || |||||||||||||||||| || |||
```


Db 961 GGATCTAACGGTATTGTTATTGTGGCTACTACCAGAACAGTTACAGACAGTACTACCGCT 1020
Qy 1021 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1080
Db 1021 GTGACCACCTTACCATTCGATCCTAACCGCGACAAAACTAAGACAATTGAAATTTTGAAA 1080
Qy 1081 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1140
Db 1081 CCTATTCCAACAACTACAATCACAACATCATATGTTGGTGTGACTACTTCCTACCTGACC 1140
Qy 1141 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1200
Db 1141 AAAACTGCACCAATTGGGGAAACTGCTACTGTTATTGTTGATATTCCATATCACACTACC 1200
Qy 1201 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCA 1260
Db 1201 ACTACTGTTACCAGTAAATGGACAGGAACAATTACTTCCACCACAACACATACTAATCCA 1260
Qy 1261 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320
Db 1261 ACTGACTCAATAGACACTGTCATTGTACAAGTTCCACTGCCAAACCCAACTGTTACTACC 1320
Qy 1321 ACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGT 1380
Db 1321 ACTGAATATTGGTCTCAATCATTTGCTACCACCACCACCATTACTGGACCACCAGGAAAC 1380
Qy 1381 ACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATATTGG 1440
Db 1381 ACTGATACTGTTTTAATCAGAGAACCACCGAACCATACTGTCACTACAACCGAGTACTGG 1440
Qy 1441 TCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTA 1500
Db 1441 TCAGAATCTTACACTACTACTAGTACTTTCACTGCTCCTCCAGGTGGAACTGATTCAGTT 1500
Qy 1501 ATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTT 1560
Db 1501 ATTATCAAGGAACCTCCAAATCCAACTGTCACAACTACCGAGTACTGGTCAGAATCTTAC 1560
Qy 1561 GCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAA 1620
Db 1561 ACTACCACTAGTACCTTCACTGCTCCTCCAGGTGGAACTGATTCAGTTATTATCAAGGAA 1620
Qy 1621 CCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACT 1680
Db 1621 CCACCAAACCATACTGTAACCACAACTGAATATTGGTCACAATCTTACACTACCACTACT 1680
Qy 1681 ACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAACCACCAAACCAC 1740
Db 1681 ACTGTCACCGCTCCACCAGGAGGTACTGATACTGTCTTAGTCAGAGAACCACCAAACCAT 1740
Qy 1741 ACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACCACTACCACTGTAACTGCA 1800
Db 1741 ACTGTTACAACTACCGAGTACTGGTCACAATCTTACACTACAACCACCACTGTTATTGCC 1800
Qy 1801 CCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCACCAAACCACACTGTCACTACT 1860
Db 1801 CCACCAGGTGGCACTGATTCGGTTATCATTAGAGAACCTCCAAATCCAACTGTCACAACC 1860
Qy 1861 ACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGC 1920
Db 1861 ACTGAGTACTGGTCTCAATCTTACGCAACTACCACTACCATTACCGCTCCTCCAGGTGAG 1920
Qy 1921 ACTGATACTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACTACTGAATACTGG 1980
Db 1921 ACCGATACTGTCCTTATTAGAGAACCACCAAACCATACTGTAACCACAACTGAGTATTGG 1980
Qy 1981 TCTCAATCATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGTT 2040
Db 1981 TCTCAATCTTATGCAACTACTACTACAATCACTGCTCCTCCAGGTGAAACCGATACCGTT 2040
Qy 2041 CTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATAT 2100
Db 2041 CTTATTAGGGAACCACCAAATCACACTGTCACTACTACTGAATACTGGTCACAATCATTT 2100
Qy 2101 GCTACAACCACCACTGTTACTGCACCACCTGGTGAAACCGATACCGTTCTTATCAGAGAG 2160
Db 2101 GCTACAACCACAACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAA 2160
Qy 2161 CCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACC 2220
Db 2161 CCACCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCTTACGCAACCACTACT 2220
Qy 2221 ACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGAGCCACCAAATCCA 2280
Db 2221 ACCATTACCGCTCCACCTGGTGAGACCGATACCGTTCTTATTAGGGAACCACCAAATCAC 2280
Qy 2281 ACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACAGTTACTGCT 2340
Db 2281 ACAGTTACTACTACTGAATACTGGTCACAATCATATGCAACTACTACCACTATCATCGCA 2340
Qy 2341 CCTCCAGGTGGTACTGACACTGTGATTATCTATGA---AAGCATGTCAAGTTCAAAGATT 2397
Db 2341 CCACCAGGTGAAACTGATACTGTTTTAATCAGAGAGCCACCAAACCCAACTGTTACCACC 2400
Qy 2398 TCTACATCCTCCAATGATATAACCAGTATCATTCCATCATTTTCCCGTCCTCATTATGTC 2457
Db 2401 ACCGAATACTGGTCTCAATCCTATACCACTGCTACTACCGTTACTGCACCACCAGGTGGA 2460
Qy 2458 AACAGCACAACCTCCGATTTGTCAACATTTGAATCTTCATCCATGAATAC---TCCTACT 2514
Db 2461 ACTGATACTGTGATTATTTATGACACCATGTCAAGTTCAGAAATTTCTTCATTTTCTCGT 2520
Qy 2515 TCTATCAGTAGTGATGGTATGTTGTTGTCTTCTACAACTTTGGTTACTGAATCAGAAACA 2574
Db 2521 CCTCATTACACCAACCATACAACTTTGTGGTCTACAACTTGGGTTATTGAAACAAAAACA 2580
Qy 2575 ACTACAGAACTGATTTGCAGTGATGGTAAAGAGTGTTCTAGATTGTCCAGTTCTTCTGGT 2634
Db 2581 ATTACAGAAACTAGCTGTGAAGGTGATAAAGGTTGTTCTTGGGTTTCTGTTTCTACTCGT 2640
Qy 2635 ATTGTCACAAATCCAGATAGCAATGAATCCTCAATCGTAACTAGTACTGTTCCTACTGCA 2694
Db 2641 ATTGTCACAATTCCTAATAATATCGAAACTCCTATGGTTACTAATACTGTTGATTCTACA 2700
Qy 2695 AGTACAATGTCTGATTCACTTTCTTCAACTGATGGTATTAGTGCTACATCTTCTGATAAT 2754
Db 2701 ACCACAGAATCCACTTCACAATCCCCATC---TGGTATTTTTTCAGAGTCAGGAGTATCT 2757
Qy 2755 GTTTCAAAATCAGGAGTATCAGTTACAACCGAAACTTCTGTTACAACTATTCAAACTACT 2814
Db 2758 GTTGAAACAGAATCTTCTACTGTTACTACTGCTCAAACAAATCCAAGTGTTCCAACAACT 2817
Qy 2815 CCAAACCCATTATCATCTTCAGTGACATCATTGACTCAGTTGTCTTCAATTCCAAGTGTT 2874
Db 2818 GAAAGTGAGGTTGTATTTACTACTAAAGGAAACAACGAAAATGGTCCTTATGAATCACCA 2877
Qy 2875 TCAGAAAGTGAAAGTAAAGTTACATTTACAAGCAATGGAGACAACCAAAGTGGTACTCAT 2934
Db 2878 TCTACTAATGTGAAATCAAGTATGGATGAAAACTCTGAATTTA---CTACTTCCACAGCT 2934
Qy 2935 GATTCACAATCTACTTCCACTGAAATTGAA---ATTGTAACAACCAGTTCTACTAAAGTT 2991
Db 2935 GCTTCCACTTCTACTGATATTGAAAATGAAACCATAGCAACAACCGGTTCCGTGGAAGCT 2994
Qy 2992 TTACCACCTGTCGTTTCTTCTAATACTGATTTGACTAGTGAACCAACAAATACCAGAGAA 3051
Db 2995 TCATCGCCTATCATTTCTTCTAGTGCTGATGAAACTACTACTGTTACTACTACTGCTGAA 3054
Qy 3052 CAACCAACTACATTATCAACTACTTCAAACTCCATCACTGAAGATATCACCACATCTCAA 3111
Db 3055 TCAACCAGTGTCATTGAACAACCAACCAATAATAATGGTGGTGGTAAAGCCCCATCTGCA 3114
Qy 3112 CCTACAGGTGATAATGGAGACAATACTTCATCAACCAATCCAGTTCCAACTGTGGCAACA 3171
Db 3115 ACTTCATCTCCATCTACAACTACAACTGCTAATAATGACTCTGTTATTACTGGTACAACA 3174

```
Qy       3172 AGTACTTTAGCATCTGCAAGTGAAGAAGACAACAAAAGCGGTTCTCATGAATCAGCAT 3229
Db       3175 TCAACCAACCAATCTCAATCTCAATCTCAATATAATTCTGATACCCAACAAACTACAT 3232

RESULT 3
AF051313
LOCUS       AF051313               4383 bp    DNA     linear   PLN 01-AUG-2000
DEFINITION  Candida albicans agglutinin-like cell surface protein (ALS8) gene,
            complete cds.
ACCESSION   AF051313
VERSION     AF051313.2  GI:9625353
KEYWORDS    .
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 4383)
  AUTHORS   Leng,P., Lee,P.R., Wishart,J.A., Wu,H. and Brown,A.J.P.
  TITLE     Sequence of the hypha-specific, agglutinin-like cell surface
            protein, ALS8 from Candida albicans
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4383)
  AUTHORS   Leng,P., Lee,P.R., Wishart,J.A., Wu,H. and Brown,A.J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1999) Molecular and Cell Biology, University of
            Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen
            AB25 2ZD, UK
COMMENT     On Aug 1, 2000 this sequence version replaced gi:4105850.
FEATURES             Location/Qualifiers
     source          1. .4383
                     /organism="Candida albicans"
                     /strain="ATCC10261"
                     /db_xref="ATCC:10261"
                     /db_xref="taxon:5476"
     gene            566. .>3840
                     /gene="ALS8"
     mRNA            566. .>3840
                     /gene="ALS8"
                     /product="agglutinin-like cell surface protein"
                     /note="minor form"
     protein_bind    574. .579
                     /gene="ALS8"
                     /note="E-box"
                     /bound_moiety="Efg1p"
     TATA_signal     599. .603
                     /gene="ALS8"
     mRNA            612. .>3840
                     /gene="ALS8"
                     /product="agglutinin-like cell surface protein"
                     /note="minor form"
     mRNA            620. .>3840
                     /gene="ALS8"
                     /product="agglutinin-like cell surface protein"
                     /note="major form"
     mRNA            630. .>3840
                     /gene="ALS8"
                     /product="agglutinin-like cell surface protein"
                     /note="minor form"
     CDS             697. .3840
                     /gene="ALS8"
                     /function="involved in cell-cell adhesion"
                     /note="hypha-specific cell surface glycoprotein; similar
                     to Candida albicans strain 1161 agglutinin-like protein
                     encoded by ALS3 gene sequence presented in GenBank
                     Accession Number U87956; contains two fewer internal
                     repeated units than the Candida albicans strain 1161 ALS3
                     gene"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like cell surface protein"
                     /protein_id="AAD02580.1"
                     /db_xref="GI:4105851"
                     /translation="MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYHYKGPG
                     TPTWNAVLGWSLDGTSASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGE
                     EFMTFSTLTCTVSNTLTPSIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTF
                     NDGGKKISINVDFERSNVDPKGYLTDSRVIPSLNKVSTLFVAPQCANGYTSGTMGFAN
                     TYGDVQIDCSNIHVGITKGLNDWNYPVSSESFSYTKTCSSNGIFITYKNVPAGYRPFV
                     DAYISATDVNSYTLSYANEYTCAGGYWQRAPFTLRWTGYRNSDAGSNGIVIVATTRTV
                     TDSTTAVTTLPFDPNRDKTKTIEILKPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDIPYHTTTTVTSKWTGTITSTTTHTNPTDSIDTVIVQVPSPNPTVTTTEYWSQSFAT
                     TTTITGPPGNTDTVLIREPPNHTVTTTEYWSESYTTTSTFTAPPGGTDSVIIKEPPNP
                     TVTTTEYWSESYTTTSTFTAPPGGTDSVIIKEPPNPTVTTTEYWSQSYTTTTTVTAPP
                     GGTDTVLVREPPNHTVTTTEYWSQSYTTTTTVIAPPGGTDTVIIREPPNHTVTTTEYW
                     SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTITAPPGETDTVLI
                     REPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVLIREPPNHTVTTTEYWSQSYATTT
                     TVTAPPGGTDTVIIYDTMSSSEISSFSRPHYTNHTTLWSTTWVIETKTITETSCEGDK
                     GCSWVSVSTRIVTIPNNIETPMVTNTVDSTTTESTSQSPSGIFSESGVSVETESSTVT
                     TAQTNPSVPTTESEVEFTTKGNNGNGPYESPSTHVKSSMDENSEFTTSTAASTSTDIE
                     NATIATTGSVEASSPIISSSADETTTITTTAESTSVIEQPTNNNGGGKAPSATSSPST
                     TTTANNDSVITGTTSTNQSQSQSQSNSDTQQTTLSQQMTSSLVSLHMLTTFDGSGSVI
                     QHSTWLCGLITLLSLFI"
BASE COUNT     1359 a    922 c    705 g   1397 t
ORIGIN

  Query Match             45.1%;  Score 1708.6;  DB 8;  Length 4383;
  Best Local Similarity   84.6%;  Pred. No. 1.1e-298;
  Matches 1918;  Conservative    0;  Mismatches  349;  Indels    0;  Gaps    0;

Qy          1 ATGCTTCAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATC 60
Db        697 ATGCTACAACAATATACATTGTTACTCATATATTTGTCGGTTGCGACTGCAAAGACAATC 756

Qy         61 ACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTC 120
Db        757 ACTGGTGTTTTCAACAGTTTTAATTCATTGACTTGGTCTAATGCTGCTACTTATCATTAT 816

Qy        121 AAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGT 180
Db        817 AAGGGACCAGGAACCCCAACTTGGAATGCTGTTTTGGGTTGGTCTTTAGATGGTACTAGT 876

Qy        181 GCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCA 240
Db        877 GCAAGTCCGGGAGATACATTCACATTGAATATGCCATGTGTGTTTAAATTTACTACTTCT 936

Qy        241 CAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 300
Db        937 CAAACATCTGTTGATTTGACTGCTCATGGTGTTAAATATGCTACATGTCAATTTCAGGCA 996

Qy        301 GGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 360
Db        997 GGTGAAGAATTTATGACCTTTTCTACATTAACATGTACTGTGAGCAATACTTTGACTCCA 1056

Qy        361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 420
Db       1057 TCTATTAAGGCTTTGGGTACTGTCACCTTACCACTTGCATTCAATGTAGGTGGAACTGGT 1116

Qy        421 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTT 480
Db       1117 TCTTCTGTTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACTAACACAGTTACATTT 1176

Qy        481 AATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCA 540
Db       1177 AATGATGGTGGCAAGAAAATCTCTATTAATGTTGATTTTGAAAGGTCAAATGTCGATCCA 1236

Qy        541 AGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTT 600
Db       1237 AAAGGGTACTTAACTGATTCCAGAGTTATACCAAGTCTCAACAAAGTGTCAACTCTTTTT 1296

Qy        601 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC 660
Db       1297 GTTGCACCACAATGTGCAAATGGTTACACATCTGGTACAATGGGATTCGCTAACACTTAT 1356

Qy        661 GGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGAT 720
Db       1357 GGTGATGTTCAAATTGACTGTTCAAATATTCATGTTGGTATTACAAAAGGATTGAATGAT 1416
```

```
Qy       721 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 780
Db      1417 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTTCATCTAATGGT 1476

Qy       781 ATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATT 840
Db      1477 ATCTTTATCACATATAAAAATGTTCCTGCCGGTTATCGTCCATTTGTTGACGCTTATATT 1536

Qy       841 TCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGC 900
Db      1537 TCTGCTACAGATGTTAACTCGTACACCTTGTCGTATGCTAATGAATATACTTGTGCTGGT 1596

Qy       901 AGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCC 960
Db      1597 GGTTATTGGCAACGTGCACCTTTCACATTAAGATGGACTGGATACAGAAATAGTGATGCT 1656

Qy       961 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1020
Db      1657 GGATCTAACGGTATTGTTATTGTGGCTACTACCAGAACAGTTACAGACAGTACTACCGCC 1716

Qy      1021 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1080
Db      1717 GTGACCACCTTACCATTCGATCCTAACCGCGACAAAACTAAGACAATTGAAATTTTGAAA 1776

Qy      1081 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1140
Db      1777 CCTATTCCAACAACTACAATCACAACATCATATGTTGGTGTGACTACTTCCTACCTGACC 1836

Qy      1141 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1200
Db      1837 AAAACTGCACCAATTGGGGAAACTGCTACTGTTATTGTTGATATTCCATATCACACTACC 1896

Qy      1201 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCA 1260
Db      1897 ACTACTGTTACCAGTAAATGGACAGGAACAATTACTTCCACCACAACACATACTAATCCA 1956

Qy      1261 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320
Db      1957 ACTGACTCAATAGACACTGTCATTGTACAAGTTCCACTGCCAAACCCAACTGTTACTACC 2016

Qy      1321 ACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGT 1380
Db      2017 ACTGAATATTGGTCTCAATCATTTGCTACCACCACCACCATTACTGGACCACCAGGGAAC 2076

Qy      1381 ACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATATTGG 1440
Db      2077 ACTGATACTGTTTTAATCAGAGAACCACCAAACCATACTGTCACTACAACCGAGTATTGG 2136

Qy      1441 TCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTA 1500
Db      2137 TCAGAATCTTACACTACCACTAGTACTTTCACTGCTCCTCCAGGTGGAACTGACTCGGTC 2196

Qy      1501 ATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTT 1560
Db      2197 ATCATCAAGGAACCTCCAAATCCAACTGTTACAACTACCGAGTACTGGTCAGAATCTTAC 2256

Qy      1561 GCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAA 1620
Db      2257 ACTACCACTAGTACCTTCACTGCTCCTCCAGGTGGAACTGACTCGGTCATCATCAAGGAA 2316

Qy      1621 CCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACT 1680
Db      2317 CCACCAAATCCAACTGTCACTACAACAGAGTACTGGTCACAATCTTACACTACCACTACT 2376

Qy      1681 ACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAACCACCAAACCAC 1740
Db      2377 ACTGTCACCGCTCCACCAGGAGGTACTGATACTGTCTTAGTCAGAGAGCCACCAAACCAT 2436

Qy      1741 ACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACCACTACCACTGTAACTGCA 1800
Db      2437 ACTGTTACAACTACCGAGTACTGGTCACAATCTTACACTACAACCACCACTGTTATTGCC 2496

Qy      1801 CCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCACCAAACCACACTGTCACTACT 1860
Db      2497 CCACCAGGTGGTACTGACACTGTTATCATTAGAGAACCACCAAACCACACTGTCACTACT 2556

Qy      1861 ACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGC 1920
Db      2557 ACTGAGTATTGGTCTCAATCTTACGCAACCACTACTACCATTACCGCTCCACCTGGTGAG 2616

Qy      1921 ACTGATACTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACTACTGAATACTGG 1980
Db      2617 ACCGATACTGTCCTTATTAGAGAACCACCAAACCATACTGTAACCACAACTGAGTATTGG 2676

Qy      1981 TCTCAATCATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGTT 2040
Db      2677 TCTCAATCCTATGCAACTACTACTACAATCACTGCTCCTCCAGGTGAAACCGATACCGTT 2736

Qy      2041 CTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATAT 2100
Db      2737 CTTATTAGGGAACCACCAAATCACACTGTCACTACTACTGAATACTGGTCACAATCATAT 2796

Qy      2101 GCTACAACCACCACTGTTACTGCACCACCTGGTGAAACCGATACCGTTCTTATCAGAGAG 2160
Db      2797 GCTACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTCTTATCAGAGAG 2856

Qy      2161 CCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACC 2220
Db      2857 CCACCAAACCACACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACC 2916

Qy      2221 ACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGA 2267
Db      2917 ACTGTTACTGCACCACCAGGAGGTACCGATACTGTGATTATTTATGA 2963
```

```
RESULT 4
AF272027
LOCUS       AF272027               4569 bp    mRNA    linear   PLN 24-OCT-2000
DEFINITION  Candida albicans agglutinin-like protein mRNA, partial cds.
ACCESSION   AF272027
VERSION     AF272027.1  GI:10952735
KEYWORDS    .
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 4569)
  AUTHORS   Chen,X. and Chen,J.-Y.
  TITLE     ALS4 (agglutinin-like sequence) of Candida albicans
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4569)
  AUTHORS   Chen,X. and Chen,J.-Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2000) Shanghai Institute of Biochemistry, Yue
            Yang Road 320, Shanghai 200031, China
FEATURES             Location/Qualifiers
     source          1. .4569
                     /organism="Candida albicans"
                     /strain="SC5314"
                     /db_xref="taxon:5476"
     CDS             1. .>4569
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAG25054.1"
                     /db_xref="GI:10952736"
                     /translation="MLLQFLLLSLCVSVATAKVITGIFDSFNSLTWTNAASYSYRGPA
                     NPTWTAVIGWSLDGATASAGDTFTLDMPCVFKFITDQTSIDLVADGRTYATCNLNSAE
                     EFTTFSSVSCTVTTTMTADTKAIGTVTLPFSFSVGGSGSDVDLANSQCFTAGINTVTF
                     NDGDTSISATVDFEKSTVASSDRILLSRILPSLSQAVSLFLPQECANGYTSGTMGFST
                     AGTGATIDCSTVHVGISNGLNDWNYPISSESFSYTKTCTSTSVLVTYQNVPAGYRPFV
                     DAYVSATRVSSYAMRYTNIYACVGAASVDDSFTHTWSGYSNSQAGSNGITIVVTTRTV
                     TDSTTAVTTLPFNSESDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPSPNPTVTTTEYWSQSYAT
                     TTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNP
```

TVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPP
GGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYW
SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVII
REPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYVTTS
TITAPPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTV
TTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTAPPGG
TDTVIIREPPSPTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQ
SYATTTTVTSPPGGTDIVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIRE
PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTV
TAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTT
TEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSYATTLTITAPPGGTN
SVIIRVHSSTNDESSESTFSTLSVPSFSGSISIVSTVSRPHYVNSTVTHLPSSSSKPV
DIPSSDVVPSTNDNSLTSLTGSENGETSVAISTTFCDDENGCQTSIPQGSVVRTTAPT
TATPTPIIGDNNGSGKSKSGELSSTGSVTTNTATPDVPSTKVPSNPGAPGTGVPPPLA
PSTETQTTNNVPGSPNIPATGTTDIIRESTTVSHTVTGNGNTGVPMNPNPALTTSTSL
TGATNSATNPSHETGVNTGSRGSTNIVTPPSSATATVVIPGTDNGATTKGQDTAGGNS
NGFTATTNTQGGNNEPGNQPGTNTTGEPVGTTGTQSVESTSQPTTLSQQTTSSLISTP
LASTFDGSGSIVQHSGWLYVLLTAISIFF"

BASE COUNT 1318 a 1186 c 787 g 1278 t

ORIGIN

Query Match 40.3%; Score 1526.4; DB 8; Length 4569;
Best Local Similarity 77.7%; Pred. No. 8.8e-266;
Matches 1845; Conservative 0; Mismatches 531; Indels 0; Gaps 0;

```
Qy          1 ATGCTTCAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATC 60
              ||||||  |||||||   ||| ||  ||| | | | ||| |||| |  ||||||   ||
Db          1 ATGCTTTTACAATTTTTGTTGCTAAGCCTCTGTGTATCAGTTGCTACGGCAAAGGTTATT 60

Qy         61 ACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTC 120
              || ||  |||| |||||||||||||| ||||||||| ||||||||||  |||| | |
Db         61 ACGGGCATTTTCGATAGTTTTAATTCGTTAACTTGGACCAATGCTGCTTCTTATTCATAT 120

Qy        121 AAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGT 180
              | ||| ||||   | || |||||||  |||||  | || |||||  |||||||  | | |
Db        121 AGAGGTCCAGCTAATCCTACTTGGACCGCTGTAATAGGATGGTCTTTAGATGGAGCTACT 180

Qy        181 GCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCA 240
              || | | || || |||||||| ||  | |||||||||| || |||| || ||||
Db        181 GCTAGTGCTGGTGACACATTCACGTTAGACATGCCATGTGTTTTCAAATTTATTACTGAT 240

Qy        241 CAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 300
              ||||| ||  ||||||||  ||| ||||||  ||  ||||||||||| | ||  | |||
Db        241 CAAACGTCAATTGATTTAGTTGCTGATGGTCGTACTTATGCTACTTGTAATTTGAACTCT 300

Qy        301 GGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 360
              |  ||||| || || ||||||||||   | ||||||||||      |  |||   |
Db        301 GCCGAAGAGTTTACTACTTTTTCTAGTGTGTCATGTACTGTGACTACTACAATGACTGCT 360

Qy        361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 420
              ||  || ||  | || || || || |||||  |  ||||||  ||| || ||| |||||
Db        361 GACACCAAAGCCATAGGAACTGTAACATTACCTTTCTCATTCAGTGTGGGGGGATCAGGT 420

Qy        421 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTT 480
              |||     |||||||| | ||||| ||||||||||||| || | |||||||||  || |||
Db        421 TCAGATGTTGATTTGGCAAATTCTCAATGTTTTACTGCAGGAATCAATACAGTTACTTTT 480

Qy        481 AATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCA 540
              ||||||||||  |     || ||       ||||| |||||||| ||||||| |   |
Db        481 AATGATGGTGACACTAGCATTTCCGCAACAGTTGATTTTGAAAAATCAACCGTGGCCTCC 540

Qy        541 AGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTT 600
              || |    | | |      || ||| || | || |||||      | |    || |||||||
Db        541 AGCGATCGTATCTTGTTGTCAAGAATTTTACCCAGTCTTTCACAAGCAGTAAGTCTTTTT 600

Qy        601 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC 660
              |  | | | |||||| |||||||| || |||||||||||||| || || | |  |
Db        601 CTTCCCCAAGAATGTGCAAATGGTTATACTTCTGGTACAATGGGATTTTCAACTGCTGGT 660

Qy        661 GGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGAT 720
              ||  | | |||| ||||| || |   ||||| | || ||  |||| || || ||||||
Db        661 ACTGGTGCTACTATAGATTGTTCCACAGTTCATGTCGGGATATCAAATGGGTTGAATGAT 720

Qy        721 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 780
              ||||||||||  |||||  ||||| ||| |||||| || || |||||||| |   |
Db        721 TGGAATTATCCAATTTCACTGGAATCTTTTTCTTACACAAAGACCTGTACATCAACAAGT 780

Qy        781 ATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATT 840
              ||    | | |||||| ||||| ||||| || ||||||||||| | |||||||| ||
Db        781 GTTTTAGTAACATATCAGAATGTTCCTGCCGGATATCGTCCATTTGTCGATGCTTATGTT 840

Qy        841 TCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGC 900
              ||||| |||   || | |  |||| || |   ||| || |||   ||| |||||| |||
Db        841 TCTGCAACACGAGTCAGCTCATATGCTATGAGATACACTAATATATATGCTTGTGTTGGT 900

Qy        901 AGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCC 960
                    |         | | | || |||   |  |||   ||||| |  |||||| | ||
Db        901 GCGGCTTCTGTTGATGACTCATTTACTCATACTTGGCTGGGATATAGTAATAGTCAAGCT 960

Qy        961 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1020
              || ||||| ||||||  ||||||  | ||||||||||||||||||||||||||||||||
Db        961 GGTTCTAATGGTATTACCATTGTGGTAACAACTAGAACAGTTACAGACAGTACCACTGCT 1020

Qy       1021 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1080
              || |||||||||||||||| |      ||| ||||||||||||||||||||||| |||
Db       1021 GTGACTACTTTACCATTCAATTCCGAATCTGACAAAACCAAAACAATCGAAATTTTACAA 1080

Qy       1081 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1140
              |||||||||| |||||||| |||||||||||||||||||||| |||||||| ||||||
Db       1081 CCTATTCCAACAACTACCATTACAACTTCATATGTTGGTGTGACAACTTCCTACCTGACT 1140

Qy       1141 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1200

Qy       1201 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCA 1260
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       1201 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACTACAACTCGTACCAATCCA 1260

Qy       1261 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320
              |||||||| || || || || || || ||||||||||||||||||||||||||| | |||
Db       1261 ACTGATTCTATAGATACTGTCGTTGTGCAAGTTCCACTGCCAAATCCAACTGTCACTACA 1320

Qy       1321 ACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGT 1380
              || || |||||||||||||| | ||| || || ||||| ||||||||||||||||||||
Db       1321 ACCGAGTATTGGTCTCAGTCATATGCCACTACTACTACTGTTACTGCTCCTCCAGGTGGT 1380

Qy       1381 ACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATATTGG 1440
              || ||  | || ||||||||||||||||||| |  |||||||||| || || ||||||
Db       1381 ACTGACTCAGTAATTATCAGAGAGCCACCAAATCCAACTGTCACTACAACCGAGTATTGG 1440

Qy       1441 TCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTA 1500
              || ||||| |||||||||||||||| |||||||||||||||| || ||||||||||||
Db       1441 TCTCAATCTTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGAGGCACTGACTCAGTA 1500

Qy       1501 ATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTT 1560
              |||||||||||||| ||||| |||||||| || || || || |||||||| || ||||||
Db       1501 ATTATCAGAGAACCTCCAAACCCAACTGTAACCACCACTGAATATTGGTCACAGTCCTTT 1560

Qy       1561 GCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAA 1620
              |||||||||||||| |||||||||||||||| || ||||||||||||||||||||||||
Db       1561 GCTACTACTACTACTGTTACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAG 1620

Qy       1621 CCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACT 1680
              || ||||| |||||||||||||| || |||||||| |||||||| || || || |||
Db       1621 CCACCAAATCCAACTGTCACCACCACCGAGTATTGGTCACAATCTTATGCTACAACTACT 1680

Qy       1681 ACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAACCACCAAACCAC 1740
              ||||| |||||||||||||| || |||||||||||||||||||||||||||| ||||| |
Db       1681 ACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCTCCAAATCCA 1740

Qy       1741 ACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACCACTACCACTGTAACTGCA 1800
              ||||||| || || || || |||||||| || | ||| || ||||| ||||| |||||
Db       1741 ACTGTCACCACCACCGAGTATTGGTCACAGTCCTTTGCTACTACTACTACTGTTACTGCT 1800

Qy       1801 CCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCACCAAACCACACTGTCACTACT 1860
```

```
Db       1801 CCTCCAGGAGGCACTGACTCAGTGATTATCAGAGAACCTCCAAACCCAACTGTAACCACC 1860

Qy       1861 ACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGC 1920

Db       1861 ACTGAATATTGGTCACAGTCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGAGGC 1920

Qy       1921 ACTGATACTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACTACTGAATACTGG 1980

Db       1921 ACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACAGTTACCACCACCGAGTATTGG 1980

Qy       1981 TCTCAATCATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGTT 2040

Db       1981 TCACAATCTTATGCTACAACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTA 2040

Qy       2041 CTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATAT 2100

Db       2041 ATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACCGAGTATTGGTCACAATCTTAT 2100

Qy       2101 GCTACAACCACCACTGTTACTGCACCACCTGGTGAAACCGATACCGTTCTTATCAGAGAG 2160

Db       2101 GCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTGATTATCAGAGAA 2160

Qy       2161 CCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACC 2220

Db       2161 CCTCCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAGTCATATGTCACTACCTCA 2220

Qy       2221 ACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGAGCCACCAAATCCA 2280

Db       2221 ACTATCACAGCTCCACCGGGAGGCACTGACACTGTGATTATTAGAGAGCCACCAAATTAC 2280

Qy       2281 ACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACAGTTACTGCT 2340

Db       2281 ACGGTTACTACAACTGAATATTGGTCCCAATCTTATGCTACAACTACTACAGTCACCGCT 2340

Qy       2341 CCTCCAGGTGGTACTGACACTGTGATTATCTATGAA 2376

Db       2341 CCTCCAGGTGGTACTGACTCAGTGATTATCAGAGAA 2376
```

```
RESULT 5
AF025429
LOCUS       AF025429               4723 bp    DNA     linear   PLN 16-DEC-1997
DEFINITION  Candida albicans agglutinin-like adhesin (ALA1) gene, complete cds.
ACCESSION   AF025429
VERSION     AF025429.1  GI:2522218
KEYWORDS    .
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 4723)
  AUTHORS   Gaur,N.K. and Klotz,S.A.
  TITLE     Expression, cloning, and characterization of a Candida albicans
            gene, ALA1, that confers adherence properties upon Saccharomyces
            cerevisiae for extracellular matrix proteins
  JOURNAL   Infect. Immun. 65 (12), 5289-5294 (1997)
  MEDLINE   98053977
   PUBMED   9393828
REFERENCE   2  (bases 1 to 4723)
  AUTHORS   Gaur,N.K. and Klotz,S.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1997) Research Service, VA Medical Center, 4801
            Linwood Blvd., Kansas City, MO 64128, USA
FEATURES             Location/Qualifiers
     source          1. .4723
                     /organism="Candida albicans"
                     /db_xref="taxon:5476"
     gene            327. .4586
                     /gene="ALA1"
     CDS             327. .4586
                     /gene="ALA1"
                     /function="cell adhesion protein"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like adhesin"
                     /protein_id="AAB88883.1"
                     /db_xref="GI:2522219"
                     /translation="MIQQFTLLFLYLSFATAKAITGIFNSIDSLTWSNAGNYAFKGPG
                     YPTWNAVLGWSLDGTSANPGDTFILNMPCVFKFTASQKSVDLTADGVKYATCQFYSGE
                     EFTTFSSLKCTVNNNLRSSIKALGTVTLPIAFNVGGTGSSVDLEDSKCFTAGTNTVTF
                     NDGSKKLSIAVNFEKSTVDQSGYLTTSRFMPSLNKIATLYVAPQCENGYTSGTMGFST
                     SYGDVAIDCSNVHIGISKGVNDWNHPVTSESFSYTKSCSSFGISITYQNVPAGYRPFI
                     DAYISPSDNNQYQLSYKNDYTCVDDYWQHAPFTLKWTGYKNSDAGSNGIVIVATTRTV
                     TDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPSPNPTTTTTQFWSESFTS
                     TTTITNSLKGTDSVIVREPHNPTVTTTEFWSESFATTETITSKPEGTDSVIVREPHNP
                     TVTTTEFWSESYATTETITNGPEGTDSVIVREPHNPTVTTTKFWSESYATTETITNKP
                     EGTDSVIVKEPYNPTVTTTEFWSESYATTETITNGPEGTDSVIVREPHNPTVTTTEFW
                     SESYATTETITTGPLGTDSIVIHDPLEESSSTTAIESSDSNISSSAQESSSSVEQSSS
                     IVGLSSSSDIPLSSDMPSSSSTGLTSSESSTVSSYDSDSSSSIESSTLSSSDRCSSSI
                     SDTTSFWDSSSSDLESTSITWSSSIDAQSSHLVQSVSNSISTSQELSSSSSEESSTFA
                     TDALVSSDASSILSSDTSSYYPSSTISSSDDFPHTIAGESDSLSISFITSTVEISSDS
                     VSLTSDPASSFDSSSSLNSDSSSSPSSDQSDILTSSSFSTLVVPSFSLSSSSSLSLTY
                     PHYVNSTTYHASESESSSVASPSMASESANDDTYTLSESTDTTSSIGTDSSTVTFCRR
                     DNGDGCIVTGMPSSSIDSEQTSDVTTTSSFVASSTPTSAEQSITDNPNIDSSQTSASS
                     STKSSVSVSDTVVNSILLSETSTLSSDDSTSSDTSISSTTNSDTGNINAGSSHTSTAS
                     IKESSIQKTGVTLSSSYLSTKLSSTSDITIELITTELITTELTTIEDNEPNTFTSTPS
                     SHSEIFSSDNSVLSKQVDRESTIKTSPTTDVTTVSSLSVHSTEASTATLGENSFSNVA
                     STPSNIATSLRSTSSSSNHATESSGTVKSEASAEAIPSPPTSTDNRLSYSTEEAKGIT
                     YANSGSTNNLITESQVAAPTDSTSVLIENPVVTSTFDDNSSAAVDQPSKTKSIEESIM
                     NPDSTNETNNGFIATLSQAQVPSSSIHSELISTTTAKTTDASMNGDSAASNSQPTTLI
                     QQVATSSYNQPLITTYAGSSSATKHPSWLLKFISVALFFFL"
BASE COUNT     1441 a   1013 c    803 g   1466 t
ORIGIN
```

```
  Query Match             36.8%;  Score 1392.8;  DB 8;  Length 4723;
  Best Local Similarity   81.6%;  Pred. No. 1.1e-241;
  Matches 1610;  Conservative    0;  Mismatches  362;  Indels    0;  Gaps    0;

Qy          1 ATGCTTCAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATC 60

Db        327 ATGATTCAACAATTTACATTGTTATTCCTATATTTGTCGTTTGCGACTGCAAAGGCGATC 386

Qy         61 ACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTC 120

Db        387 ACTGGTATTTTCAATAGTATTGACTCATTAACTTGGTCCAATGCTGGCAATTACGCTTTC 446

Qy        121 AAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGT 180

Db        447 AAAGGACCAGGATACCCAACTTGGAATGCTGTGTTGGGTTGGTCATTAGATGGTACCAGT 506

Qy        181 GCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCA 240

Db        507 GCCAATCCAGGGGATACATTCATATTAAACATGCCATGTGTGTTTAAATTCACTGCTTCC 566

Qy        241 CAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 300

Db        567 CAAAAATCTGTTGATTTGACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 626

Qy        301 GGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 360

Db        627 GGTGAAGAGTTTACAACTTTTTCTTCATTAAAATGTACAGTGAACAATAATTTGAGATCA 686

Qy        361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 420

Db        687 TCTATTAAGGCTTTGGGTACGGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGG 746

Qy        421 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTT 480

Db        747 TCATCAGTTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACGGTAACATTT 806

Qy        481 AATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCA 540

Db        807 AATGATGGCAGTAAAAAGCTCTCAATTGCTGTTAATTTTGAAAAGTCAACAGTTGATCAA 866
```

```
Qy       541 AGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTT 600
Db       867 AGTGGGTATTTGACTACTTCCAGATTTATGCCGAGTCTCAATAAAATTGCTACTCTTTAT 926

Qy       601 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC 660
Db       927 GTGGCACCACAATGTGAAAACGGTTACACATCTGGTACAATGGGATTCTCCACTAGTTAT 986

Qy       661 GGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGAT 720
Db       987 GGGGATGTTGCTATTGACTGTTCAAATGTACATATTGGTATTTCAAAAGGAGTAAATGAT 1046

Qy       721 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 780
Db      1047 TGGAATCATCCAGTTACGTCTGAATCATTTAGTTACACTAAAAGTTGTTCATCTTTTGGT 1106

Qy       781 ATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATT 840
Db      1107 ATCTCTATCACATATCAAAATGTTCCTGCCGGTTATCGTCCATTTATTGACGCTTATATT 1166

Qy       841 TCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGC 900
Db      1167 TCTCCCTCAGATAATAACCAGTATCAATTGTCGTATAAAAATGACTATACTTGTGTTGAT 1226

Qy       901 AGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCC 960
Db      1227 GATTATTGGCAGCATGCACCTTTCACTTTAAAATGGACTGGATATAAGAATAGTGATGCC 1286

Qy       961 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1020
Db      1287 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1346

Qy      1021 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1080
Db      1347 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1406

Qy      1081 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1140
Db      1407 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1466

Qy      1141 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1200
Db      1467 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1526

Qy      1201 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCA 1260
Db      1527 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACTACAACTCGTACCAATCCA 1586

Qy      1261 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320
Db      1587 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTACAACTACA 1646

Qy      1321 ACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGT 1380
Db      1647 ACCCAGTTTTGGTCAGAGTCATTTACTAGTACTACTACAATCACCAACTCTTTAAAAGGT 1706

Qy      1381 ACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATATTGG 1440
Db      1707 ACAGACTCAGTCATTGTTAGAGAACCACACAACCCTACTGTTACCACCACGGAGTTTTGG 1766

Qy      1441 TCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTA 1500
Db      1767 TCAGAGTCATTTGCCACTACTGAGACAATCACCAGCAAGCCAGAAGGCACAGACTCAGTC 1826

Qy      1501 ATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTT 1560
Db      1827 ATTGTTAGAGAACCACACAACCCTACTGTGACTACCACCGAGTTTTGGTCAGAGTCGTAT 1886

Qy      1561 GCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAA 1620
Db      1887 GCTACTACTGAGACCATCACCAATGGTCCAGAAGGAACAGACTCAGTCATTGTTAGAGAA 1946

Qy      1621 CCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACT 1680
Db      1947 CCACACAACCCTACTGTGACTACCACCAAGTTTTGGTCAGAGTCGTATGCTACTACTGAG 2006

Qy      1681 ACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAACCACCAAACCAC 1740
Db      2007 ACCATCACCAACAAGCCAGAAGGCACAGACTCAGTCATTGTCAAGGAACCATACAACCCT 2066

Qy      1741 ACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACCACTACCACTGTAACTGCA 1800
Db      2067 ACTGTGACTACCACCGAGTTTTGGTCAGAGTCGTATGCTACTACTGAGACCATCACCAAT 2126

Qy      1801 CCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCACCAAACCACACTGTCACTACT 1860
Db      2127 GGTCCAGAAGGAACAGACTCAGTCATTGTTAGAGAACCACACAACCCTACTGTGACTACC 2186

Qy      1861 ACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGC 1920
Db      2187 ACCGAGTTTTGGTCAGAGTCGTATGCTACTACTGAAACAATTACTACAGGGCCACTTGGT 2246

Qy      1921 ACTGATACTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACTACTG 1972
Db      2247 ACTGATAGTATCGTTATACATGATCCATTGGAAGAACTGTCTTCTACTACTG 2298
```

RESULT 6
AF068866

```
LOCUS       AF068866               3813 bp    DNA     linear   PLN 29-JAN-2001
DEFINITION  Candida albicans agglutinin-like protein (ALS5) gene, ALS5-1
            allele, complete cds.
ACCESSION   AF068866
VERSION     AF068866.1  GI:4903268
KEYWORDS    .
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 3813)
  AUTHORS   Hoyer,L.L. and Hecht,J.E.
  TITLE     The ALS5 gene of Candida albicans and analysis of the Als5p
            N-terminal domain
  JOURNAL   Yeast 18 (1), 49-60 (2001)
  MEDLINE   21064501
   PUBMED   11124701
REFERENCE   2  (bases 1 to 3813)
  AUTHORS   Hoyer,L.L., Ho,M. and Hecht,J.E.
  TITLE     The ALS5, ALS6 and ALS7 genes of Candida albicans
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3813)
  AUTHORS   Hoyer,L.L., Ho,M. and Hecht,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) Veterinary Pathobiology, University of
            Illinois at Urbana-Champaign, 2001 S. Lincoln Avenue, Urbana, IL
            61802, USA
FEATURES             Location/Qualifiers
     source          1. .3813
                     /organism="Candida albicans"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="3"
                     /note="SfiI fragment P"
     gene            1. .3813
                     /gene="ALS5"
                     /allele="ALS5-1"
     CDS             1. .3813
                     /gene="ALS5"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAD32849.1"
                     /db_xref="GI:4903269"
                     /translation="MIQQFTLLFLYLSFATAKAITGIFNSIDSLTWSNAGNYAFKGPG
                     YPTWNAVLGWSLDGTSANPGDTFILNMPCVFKFTASQKSVDLTADGVKYATCQFYSGE
                     EFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGINTVTF
```

```
                    NDGSKKLSIAVNFEKSTVDRSGYLTTSRFMPSLNKIATLYVAPQCENGYTSGTMGFST
                    SYGDVAIDCSNVHIGISKGVNDWNHPVTSESFSYTKSCSSFGISITYQNVPAGYRPFI
                    DAYISPSDNNQYQLSYKNDYTCVDDYWQHAPFTLKWTGYKNSDAGSNGIVIVATTRTV
                    TDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATLI
                    VDVPYHTTTTVTSKWTGTITMTTTRTNPTDSIDTVVVQVPLPNPTTTTTQFWSESFTS
                    TTTITNKPEGTDSVIVKEPHNPTVTTTEFWSESYATTETITTGPLGTDSIVIHDPLEE
                    SSSTTAIESSDSNISSSAQESSSSVEQSSSIVGLSSSSDIPLSSDMPSSSSTGLTSSE
                    SSTVSSYDSDSSSSSELSTFSSSESYSSSISDTTNFWDSSSSDLESTSITWSSSIDAQ
                    SSQSVQSVSNSISTSQETTSSSGEESNTSVTDILVSSDASSILNSDISSYYPSSTISL
                    SDDFPHTIAGEPDSRSSSSIASTVEISSDLVSLTSDPTSSFDSSSSLNSDSSSPFSD
                    ESDISASSSFSTLVAPSFSLSSSSSLSLIYPHYVNSTTYHASESESSSVASPSVASES
                    ANDDTHTLSESTDTTSIIGTDSSTVTFCRRDNGDGCIVTGITSSSIDSEQTSDVTTTS
                    SFVASSTPTSAEQSITDNPNIDSSQTSASSSTKSSVSVSDTVVNSILLSETSTLSSDD
                    GTSSDTSISSTTNSDTGNINAGSSHKSTASIKESSIQKTGVTLSSSYLSTKLSSTSDI
                    TIELITTELTTIEDNEPNTFTSTPSSHSEIFSSDNSVLSKQVDRESTIKTSPTTDVTT
                    VSSLSVHSTEASTATLGENSFSNVASTPLNTATSLRSTSSSSNHATESSGTVKSEASV
                    EAIPSPPTSTDNRLSYSTEEAEGITYANSGSTNNLITESQVAAPTDSTSVLIENLVVT
                    STFDDNSSAAVDQPSKTKSIEESIMNPDSTNETNNGFIATLSQAQVPSSSIHSELIST
                    TTAKTTDASMNGDSAASNSQPTTLIQQVATSSYNQPLITTYAGSSSATKHPSWLLKFI
                    SVALFFFL"
BASE COUNT     1172 a    819 c    641 g   1181 t
ORIGIN

  Query Match           31.4%;  Score 1188.8;  DB 8;  Length 3813;
  Best Local Similarity  78.8%;  Pred. No. 7.9e-205;
  Matches 1418;  Conservative    0;  Mismatches  382;  Indels    0;  Gaps    0;

Qy          1 ATGCTTCAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATC 60
              ||| ||||||||||||||||||||||||||||||||  |||| | ||||||| | |||
Db          1 ATGATTCAACAATTTACATTGTTATTCCTATATTTGTCGTTTGCGACTGCAAAGGCGATC 60

Qy         61 ACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTC 120
              |||||| ||||  ||||| || | |||||||||||||||||||||  ||||| ||||||
Db         61 ACTGGTATTTTCAATAGTATTGACTCATTAACTTGGTCCAATGCTGGCAATTACGCTTTC 120

Qy        121 AAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGT 180
              ||||| |||||||||||||||||||||||| |||||||||| ||||||||||||||||
Db        121 AAAGGACCAGGATACCCAACTTGGAATGCTGTGTTGGGTTGGTCATTAGATGGTACCAGT 180

Qy        181 GCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCA 240
              |||||||||| |||||||||| ||| || |||||||||||||||||  ||| ||||
Db        181 GCCAATCCAGGAGATACATTCATATTAAACATGCCATGTGTGTTTAAATTCACTGCTTCC 240

Qy        241 CAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 300
              |||| |||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        241 CAAAAATCTGTTGATTTGACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 300

Qy        301 GGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 360
              |||||||| || ||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGTGAAGAGTTTACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 360

Qy        361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 420

Qy        421 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTT 480
              |||||||||||||||||||||||||||||||||||||||||| |||||| || ||||||
Db        421 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTATCAATACGGTAACATTT 480

Qy        481 AATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCA 540
              ||||||||   |||| |  |||||||| |||| | ||||||||||||| ||||||| |
Db        481 AATGATGGCAGTAAAAAGCTCTCAATTGCTGTTAATTTTGAAAAGTCAACAGTTGATCGA 540

Qy        541 AGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTT 600
              ||||  ||||||  | |||||||| ||||||| ||||||||||||  |  | ||||||| |
Db        541 AGTGGGTATTTGACTACTTCCAGATTTATGCCGAGTCTCAATAAAATTGCTACTCTTTAT 600

Qy        601 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC 660
              |||||||||||||||||||| |||||||||||||||||||||||| ||||||| |||| |
Db        601 GTGGCACCACAATGTGAAAACGGTTACACATCTGGTACAATGGGATTCTCCACTAGTTAT 660

Qy        661 GGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGAT 720
              || || |||||||||| || |||||| | |||||||||||  |||||||| ||||||||

Db        661 GGGGATGTTGCTATTGACTGTTCAAATGTACATATTGGTATTTCAAAAGGAGTAAATGAT 720

Qy        721 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 780
              |||||| |||| ||| | |||||||||||||||||||||||||  ||| |||||  |||
Db        721 TGGAATCATCCAGTTACGTCTGAATCATTTAGTTACACTAAAAGCTGTTCATCTTTTGGT 780

Qy        781 ATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATT 840
              ||    || | ||||||||||||| ||||| |||||||||||||||||||| |||||||||
Db        781 ATCTCTATCACATATCAAAATGTTCCTGCCGGTTATCGTCCATTTATTGACGCTTATATT 840

Qy        841 TCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGC 900
              ||| |  ||||| |||||| |||   ||  | |||| ||||| |||||||||| ||
Db        841 TCTCCCTCAGATAATAACCAGTATCAATTGTCGTATAAAAATGACTATACTTGTGTTGAT 900

Qy        901 AGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCC 960
                | |  |||| |  |||||||||||||| |||||||||||| ||||||||||||||||
Db        901 GATTATTGGCAACATGCACCTTTCACTTTAAAATGGACTGGATATAAGAATAGTGATGCC 960

Qy        961 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1020

Qy       1021 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1080

Qy       1081 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1140

Qy       1141 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1200
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1141 AAGACTGCACCAATTGGTGAAACAGCTACTCTTATTGTTGATGTGCCATATCATACTACC 1200

Qy       1201 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCA 1260
              |||||||||| ||| |||||||||||||||| || |  || ||||||||||||||||||
Db       1201 ACAACTGTTACAAGTAAATGGACAGGAACAATTACGATGACTACAACTCGTACCAATCCA 1260

Qy       1261 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320
              |||||||||||||||||||||||||||||||||||||||| ||||||||||| | |||
Db       1261 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTTCCAAATCCAACTACAACTACA 1320

Qy       1321 ACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGT 1380
              || | | |||||  |||| ||| |||  || |||||  | ||      |||| ||
Db       1321 ACCCAGTTTTGGTCAGAGTCATTTACTAGTACTACTACAATCACCAACAAGCCAGAAGGC 1380

Qy       1381 ACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATATTGG 1440
              || ||  | || ||| |||  || ||||  || | |||||| || || || || | ||||
Db       1381 ACAGACTCAGTCATTGTCAAGGAACCACACAATCCTACTGTTACCACCACAGAGTTTTGG 1440

Qy       1441 TCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTA 1500
              ||| |||| | ||| ||||||   ||  ||||| |   |||  ||||||||      |
Db       1441 TCAGAATCATATGCCACTACTGAAACAATTACTACAGGGCCACTTGGTACTGATAGTATC 1500

Qy       1501 ATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTT 1560
              ||||    || |||    |  |       |||| || |    |  |||       | |
Db       1501 GTTATACATGATCCATTGGAAGAACTGTCTTCTACTACTGCTATTGAGTCGAGTGATTCT 1560

Qy       1561 GCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAA 1620
                || | |  ||| |   |  | |  ||     |||      |   | | ||
Db       1561 AATATTTCAAGCTCAGCTCAAGAATCATCCAGTCTGGTTGAGCAGTCATCAAGTATAGTT 1620

Qy       1621 CCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACT 1680
                    ||| |      |  |  |  ||    |   | |   | | | |||||
Db       1621 GGATTGTCATCAAGTTCAGATATACCATTAAGTTCAGACATGCCATCATCGAGCTCAACT 1680

Qy       1681 ACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAACCACCAAACCAC 1740
                | ||  ||  |          |||| ||||      |   || ||  || |||
Db       1681 GGGTTAACATCTAGTGAGTCGTCTACTGTCTCAAGTTATGATAGCGATTCATCAAGTAGT 1740

Qy       1741 ACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACCACTACCACTGTAACTGCA 1800
              | ||      ||||    |  | ||  |   ||  |  | | || | |||  ||  ||
Db       1741 AGTGAGTTATCTACATTTTCTAGTTCTGAGAGCTACTCGTCAAGTATCTCTGATACCACA 1800
```

```
RESULT 7
AR044076
LOCUS       AR044076               1071 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1 from patent US 5817466.
ACCESSION   AR044076
VERSION     AR044076.1  GI:5965541
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1071)
  AUTHORS   Hoyer,L.L., Livi,G.P. and Shatzman,A.R.
  TITLE     Conserved yeast nucleic acid sequences
  JOURNAL   Patent: US 5817466-A 1 06-OCT-1998;
FEATURES             Location/Qualifiers
     source          1. .1071
                     /organism="unknown"
BASE COUNT      317 a    322 c    159 g    273 t
ORIGIN

  Query Match           28.2%;  Score 1069.4;  DB 6;  Length 1071;
  Best Local Similarity   99.9%;  Pred. No. 3.8e-183;
  Matches 1070;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1300 CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACA 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACA 60

Qy       1360 GTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACT 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACT 120

Qy       1420 GTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCT 1479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCT 180

Qy       1480 CCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACC 1539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACC 240

Qy       1540 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACT 1599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACT 300

Qy       1600 GACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCC 1659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCC 360

Qy       1660 CAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATT 1719
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATT 420

Qy       1720 ATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCC 1779
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCC 480

Qy       1780 ACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCA 1839
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCA 540

Qy       1840 CCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACT 1899
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACT 600

Qy       1900 GTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCAACT 1959
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCAACT 660

Qy       1960 GTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCTCCA 2019
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCTCCA 720

Qy       2020 CCTGGTGAAACTGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACT 2079
             ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        721 CCTGGTGAAACTGATACCGTTCTCATCAGAGAGCCACCAAACCATACTGTCACTACTACT 780

Qy       2080 GAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACC 2139
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACC 840

Qy       2140 GATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCT 2199
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCT 900

Qy       2200 CAATCATATGCTACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATC 2259
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CAATCATATGCTACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATC 960

Qy       2260 ATTAGAGAGCCACCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCC 2319
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATTAGAGAGCCACCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCC 1020

Qy       2320 ACAACCACCACAGTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 2370
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCACCACAGTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 1071


RESULT 8
I65632
LOCUS       I65632                 1071 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 1 from patent US 5668263.
ACCESSION   I65632
VERSION     I65632.1  GI:2482202
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1071)
  AUTHORS   Hoyer,L.L., Livi,G.P. and Shatzman,A.R.
  TITLE     Conserved yeast nucleic acid sequences
  JOURNAL   Patent: US 5668263-A 1 16-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .1071
                     /organism="unknown"
BASE COUNT      317 a    322 c    159 g    273 t
ORIGIN

  Query Match           28.2%;  Score 1069.4;  DB 6;  Length 1071;
  Best Local Similarity   99.9%;  Pred. No. 3.8e-183;
  Matches 1070;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1300 CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACA 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACA 60

Qy       1360 GTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACT 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACT 120

Qy       1420 GTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCT 1479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCT 180

Qy       1480 CCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACC 1539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACC 240

Qy       1540 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACT 1599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACT 300

Qy       1600 GACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCC 1659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCC 360
```

```
Qy       1660 CAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATT 1719
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATT 420

Qy       1720 ATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCC 1779
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCC 480

Qy       1780 ACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCA 1839
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCA 540

Qy       1840 CCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACT 1899
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACT 600

Qy       1900 GTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCAACT 1959
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCAACT 660

Qy       1960 GTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCTCCA 2019
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCTCCA 720

Qy       2020 CCTGGTGAAACTGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACT 2079
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        721 CCTGGTGAAACTGATACCGTTCTCATCAGAGAGCCACCAAACCATACTGTCACTACTACT 780

Qy       2080 GAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACC 2139
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACC 840

Qy       2140 GATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCT 2199
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCT 900

Qy       2200 CAATCATATGCTACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATC 2259
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CAATCATATGCTACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATC 960

Qy       2260 ATTAGAGAGCCACCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCC 2319
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATTAGAGAGCCACCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCC 1020

Qy       2320 ACAACCACCACAGTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 2370
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCACCACAGTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 1071
```

```
RESULT 9
CANALS2S1
LOCUS       CANALS2S1               1404 bp    DNA     linear   PLN 16-OCT-1998
DEFINITION  Candida albicans agglutinin-like protein (ALS2) gene, 5' partial
            cds.
ACCESSION   AF024582
VERSION     AF024582.1  GI:3598672
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 1404)
  AUTHORS   Hoyer,L.L., Payne,T.L. and Hecht,J.E.
  TITLE     Identification of Candida albicans ALS2 and ALS4 and localization
            of als proteins to the fungal cell surface
  JOURNAL   J. Bacteriol. 180 (20), 5334-5343 (1998)
  MEDLINE   98440424
   PUBMED   9765564
REFERENCE   2  (bases 1 to 1404)
  AUTHORS   Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1997) Veterinary Pathobiology, University of
            Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES             Location/Qualifiers
     source          1. .1404
                     /organism="Candida albicans"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="6"
                     /map="6C"
     gene            1. .>1404
                     /gene="ALS2"
                     /note="agglutinin-like sequence"
                     /allele="ALS2-2"
     CDS             1. .>1404
                     /gene="ALS2"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAC64237.1"
                     /db_xref="GI:3598675"
                     /translation="MLLQFLLLSLCVSVATAKVITGVFNSFDSLTWTRAGNYAYKGPN
                     RPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKFITDQTSVDLTAEGVKYATCQFYSGE
                     EFTTFSSLKCTVSNTLTSSIKALGTVTLPISFNVGGTGSSVDLESSQCFKAGTNTVTF
                     NDGDKKISIDVDFEKTNEDASGFFIASRLIPSINKVSITYVAPQCANGYTSGAMGFIV
                     LTGDTTIDCSNVHVGITKGLNDWNFPVSSDSLSYNKTCSSTGISITYENVPAGYRPFF
                     DVYTSVSGQNRQLRYTNDYACVGSSLQSKPFNLRLRGYNNSEANSNGFVIVATTRTVT
                     DSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVIV
                     DVPYHTTTTVTSEWTGTITATTTRTNPTDSIDTVVVQVPSPNPTVSTTEYWSQSYATT
                     TTVTAPPGGTDSVIIREP"
     repeat_region   1297. .1404
                     /note="first of multiple copies of 108bp imperfect repeat"
                     /rpt_type=tandem
BASE COUNT      421 a    266 c    255 g    462 t
ORIGIN
```

```
  Query Match             26.9%;  Score 1016.6;  DB 8;  Length 1404;
  Best Local Similarity   83.2%;  Pred. No. 1.2e-173;
  Matches 1170;  Conservative    0;  Mismatches  234;  Indels    3;  Gaps    1;

Qy          1 ATGCTTCAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATC 60
              ||||||  |||||||   ||| ||  ||| | | | ||| |||| | ||||||    ||
Db          1 ATGCTTTTACAATTTTTGTTGCTAAGCCTCTGTGTATCAGTTGCTACTGCAAAAGTTATT 60

Qy         61 ACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTC 120
              || ||||||||  ||||||| |||| || || ||| | |  |||| ||||||||||
Db         61 ACGGGTGTTTTCAATAGTTTTGATTCGTTGACATGGACAAGAGCTGGTAATTATGCTTAT 120

Qy        121 AAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGT 180
              || || |||       |||||||||||||||||||||| ||||| |||||||||| |||
Db        121 AAGGGCCCAAATAGACCAACTTGGAATGCTGTTTTGGGCTGGTCTTTAGATGGTACTAGT 180

Qy        181 GCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCA 240
              || |||||||| || |||||||||||||||||||||||||| ||||||| || |||
Db        181 GCAAATCCAGGAGACACATTCACATTGAATATGCCATGTGTTTTTAAATTTATTACCGAT 240

Qy        241 CAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 300
              |||||||||||||||| ||||| || |||||||||||||||| ||||| |||||||
Db        241 CAAACATCTGTTGATTTGACTGCTGAAGGTGTTAAATATGCTACATGTCAGTTTTATTCA 300

Qy        301 GGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 360
              |||||||||| ||||| |||||| |||||| |||||||||| | |  |||| | ||||
Db        301 GGTGAAGAATTTACAACATTTTCTTCATTAAAATGTACTGTGAGCAATACTTTAACATCA 360

Qy        361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 420
              || |||||||| || ||||| |||||||||||||||| |||| ||||||||||||||||
Db        361 TCTATTAAGGCTTTGGGTACGGTTACTTTACCAATTTCATTTAATGTTGGTGGAACAGGT 420

Qy        421 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTT 480
              |||||   ||||||||||  |||| |||||||||  |||||  ||||| ||||| || |||
Db        421 TCATCGGTTGATTTGGAAAGTTCTCAATGTTTTAAGGCTGGCACCAACACAGTTACTTTT 480

Qy        481 AATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCA 540
              |||||||||||||| | || |||||||| ||||| ||||| ||  ||| ||  ||| ||
```

```
Db        481 AATGATGGTGATAAAAAAATTTCAATTGACGTTGATTTTGAGAAAACAAACGAAGATGCA 540

Qy        541 AGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTT 600
              |||| || |||    || || ||| |||| |||||| | || || ||  |||    || |
Db        541 AGTGGATTTTTCATAGCGTCAAGACTTATTCCAAGTATTAACAAAGTTTCAATCACTTAT 600

Qy        601 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC 660
              |||||||||||||||| |||||| |||||||||||| ||||||||||     |   |
Db        601 GTGGCACCACAATGTGCAAATGGCTACACATCTGGTGCAATGGGGTTCATAGTTCTCACT 660

Qy        661 GGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGAT 720
              ||||||  | ||||||| || |||||| ||||| ||||||||||||| ||||||||||||
Db        661 GGTGACACTACTATTGACTGTTCAAATGTTCATGTTGGTATCACAAAGGGATTAAATGAT 720

Qy        721 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 780
              ||||||| |||||| || ||||| ||||| ||||||| |||||||||| ||||||  ||
Db        721 TGGAATTTTCCGGTATCGTCTGATTCATTAAGTTACAATAAAACTTGTTCATCTACAGGT 780

Qy        781 ATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATT 840
              |||   || | |||| ||||||| || ||||||||||||||||| |||| |  |||| |
Db        781 ATTTCTATCACATATGAAAATGTCCCCGCTGGTTATCGTCCATTTTTTGACGTATATACT 840

Qy        841 TCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGC 900
              |   |||     |||  | |    ||| |||||| ||||||||| | |||| |||
Db        841 CTGGTGTCAGGCCAGAACAGACA---ATTAAGATATACTAATGATTATGCCTGTGTTGGT 897

Qy        901 AGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCC 960
              |||    | |||||||| || |||| |||||||| ||  |||||||| |||||||| ||
Db        898 AGTTCCTTACAAAGTAAGCCGTTCAATTTAAGATTGAGAGGATACAATAATAGTGAAGCT 957

Qy        961 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1020
              |||||||| |||||||||||||||||  |||||||||| |||||||| ||||||
Db        958 AATTCTAACGGTTTTGTCATTGTTGCTACAACCCGAACAGTTACTGACAGTACTACTGCT 1017

Qy       1021 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1080
              |||||||||||||| || |||||||||||||| ||||||||||||||||||||||||||
Db       1018 GTCACTACTTTACCTTTTAATCCAAGTGTTGACAAAACCAAAACAATCGAAATTTTGCAA 1077

Qy       1081 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1140
              |||||||||| || || |||||||||||||||||||||||||||||||||| ||||||
Db       1078 CCTATTCCAACAACCACTATCACAACTTCATATGTTGGTGTGACTACTTCCTACCTGACT 1137

Qy       1141 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1200
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 AAAACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1197

Qy       1201 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCA 1260
              |||||||||||||||||||||||||||||||||| |||| ||||||||||||||||||
Db       1198 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTGCCACTACAACTCGTACCAATCCA 1257

Qy       1261 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320
              |||||||| || || || || || || ||||||||||||||||||||||||||||||||||
Db       1258 ACTGATTCTATAGATACTGTCGTTGTGCAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1317

Qy       1321 ACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGT 1380
              |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||
Db       1318 ACTGAATATTGGTCTCAGTCATATGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGT 1377

Qy       1381 ACCGATACTGTGATTATCAGAGAGCCA 1407
              || ||  | || ||||||||||||||
Db       1378 ACTGACTCAGTCATTATCAGAGAGCCA 1404


RESULT 10
CA1ALS1
LOCUS       CA1ALS1                 1404 bp    DNA     linear   PLN 16-OCT-1998
DEFINITION  Candida albicans agglutinin-like protein (ALS2) gene, 5' partial
            cds.
ACCESSION   AF024580
VERSION     AF024580.1  GI:3598667
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 1404)
  AUTHORS   Hoyer,L.L., Payne,T.L. and Hecht,J.E.
  TITLE     Identification of Candida albicans ALS2 and ALS4 and localization
            of als proteins to the fungal cell surface
  JOURNAL   J. Bacteriol. 180 (20), 5334-5343 (1998)
  MEDLINE   98440424
   PUBMED   9765564
REFERENCE   2  (bases 1 to 1404)
  AUTHORS   Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1997) Veterinary Pathobiology, University of
            Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES             Location/Qualifiers
     source          1. .1404
                     /organism="Candida albicans"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="6"
                     /map="6; C"
     CDS             1. .>1404
                     /gene="ALS2"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAC64235.1"
                     /db_xref="GI:3598670"
                     /translation="MLLQFLLLSLCVSVATAKVITGVFNSFDSLTWTRAGNYAYKGPN
                     RPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKFITDQTSVDLTAEGVKYATCQFYSGE
                     EFTTFSSLKCTVSNTLTSSIKALGTVTLPISFNVGGTGSSVDLESSQCFKAGTNTVTF
                     NDGDKKISIDVDFEKTNEDASGYFIASRLIPSINKVSITYVAPQCANGYTSGAMGFIV
                     LTGDTTIDCSNVHVGITKGLNDWNFPVSSDSLSYNKTCSSTGISITYENVPAGYRPFF
                     DVYTSVSGQNRQLRYTNDYACVGSSLQSKPFNLRLRGYNNSEANSNGFVIVATTRTVT
                     DSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVIV
                     DVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPSPNPTVTTTEYWSQSYATT
                     TTVTAPPGGTDSVIIREP"
     repeat_region   1297. .1404
                     /note="first of multiple copies of imperfect 108 bp
                     repeat"
                     /rpt_type=tandem
BASE COUNT      422 a    268 c    253 g    461 t
ORIGIN

  Query Match           26.6%;  Score 1006;  DB 8;  Length 1404;
  Best Local Similarity  82.7%;  Pred. No. 9.8e-172;
  Matches 1163;  Conservative    0;  Mismatches  240;  Indels    3;  Gaps    1;

Qy          1 ATGCTTCAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATC 60
              |||||| |||||||   ||| ||  ||| | | | ||| |||| | ||||||    ||
Db          1 ATGCTTTTACAATTTTTGTTGCTAAGCCTCTGTGTATCAGTTGCTACTGCAAAAGTTATT 60

Qy         61 ACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTC 120
              || ||||||||  ||||||||| |||| || || ||| | |  |||| ||||||||||
Db         61 ACGGGTGTTTTCAATAGTTTTGATTCGTTGACATGGACAAGAGCTGGTAATTATGCTTAT 120

Qy        121 AAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGT 180
              || || |||       |||||||||||||||||||||||| ||||| ||||||||||| |||
Db        121 AAGGGCCCAAATAGACCAACTTGGAATGCTGTTTTGGGCTGGTCTTTAGATGGTACTAGT 180

Qy        181 GCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCA 240
              || |||||||| || |||||||||||||||||||||||| ||||||| || |||
Db        181 GCAAATCCAGGAGACACATTCACATTGAATATGCCATGTGTTTTTAAATTTATTACCGAT 240

Qy        241 CAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 300
              |||||||||||||||| ||||| || |||||||||||||||| ||||| |||||||
Db        241 CAAACATCTGTTGATTTGACTGCTGAAGGTGTTAAATATGCTACATGTCAGTTTTATTCA 300

Qy        301 GGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 360
              |||||||||| ||||| |||||| |||||||||||| | |  |||| | ||||
Db        301 GGTGAAGAATTTACAACATTTTCTTCATTAAAATGTACTGTGAGCAATACTTTAACATCA 360
```

```
Qy        361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 420
              || |||||||| || ||||| ||||||||||||||| |||| ||||||||||||||||||
Db        361 TCTATTAAGGCTTTGGGTACGGTTACTTTACCAATTTCATTTAATGTTGGTGGAACAGGT 420

Qy        421 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTT 480
              |||||   |||||||||  |||| |||||||||  ||||| ||||| ||||| || |||
Db        421 TCATCGGTTGATTTGGAAAGTTCTCAATGTTTTAAGGCTGGCACCAACACAGTTACTTTT 480

Qy        481 AATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCA 540
              |||||||||||||| | |||||||||| ||||| ||||| ||  ||| ||  ||| ||
Db        481 AATGATGGTGATAAAAAAATCTCAATTGACGTTGATTTTGAGAAAACAAACGAAGATGCA 540

Qy        541 AGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTT 600
              |||| ||||||    || || ||| |||| |||||| | || || ||  |||     || |
Db        541 AGTGGATATTTCATAGCGTCAAGACTTATTCCAAGTATTAACAAAGTTTCAATCACTTAT 600

Qy        601 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC 660
              |||||||||||||||| |||||| |||||||||||| ||||||||||     |   |
Db        601 GTGGCACCACAATGTGCAAATGGCTACACATCTGGTGCAATGGGGTTCATAGTTCTCACT 660

Qy        661 GGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGAT 720
              ||||||  | |||||||| || |||||| ||||| ||||||| ||||| ||||||||||||
Db        661 GGTGACACTACTATTGACTGTTCAAATGTTCATGTTGGTATTACAAAGGGATTAAATGAT 720

Qy        721 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 780
              ||||||| |||||| || ||||| ||||| ||||||| |||||||||| ||||||  ||
Db        721 TGGAATTTTCCGGTATCGTCTGATTCATTAAGTTACAATAAAACTTGTTCATCTACAGGT 780

Qy        781 ATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATT 840
              |||   || | |||| |||||||| || |||||||||||||||||| |||| |  |||| |
Db        781 ATTTCTATCACATATGAAAATGTCCCCGCTGGTTATCGTCCATTTTTTGACGTATATACT 840

Qy        841 TCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGC 900
              |   |||     |||  | |    |||  |||||| ||||||||| | |||| |||
Db        841 CTGGTGTCAGGCCAGAACAGACA---ATTAAGATATACTAATGATTATGCCTGTGTTGGT 897

Qy        901 AGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCC 960
              |||    | |||||||| || |||| ||||||||  ||  |||||||| |||||||| ||
Db        898 AGTTCCTTACAAAGTAAGCCGTTCAATTTAAGATTGAGAGGATACAATAATAGTGAAGCT 957

Qy        961 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1020
                |||||||| |||||||||||||||||||  |||||||||| |||||||| ||||||
Db        958 AATTCTAACGGTTTTGTCATTGTTGCTACAACCCGAACAGTTACTGACAGTACTACTGCT 1017

Qy       1021 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1080
              |||||||||||||| || |||||||||||||| ||||||||||||||||||||||||||
Db       1018 GTCACTACTTTACCTTTTAATCCAAGTGTTGACAAAACCAAAACAATCGAAATTTTGCAA 1077

Qy       1081 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1140
              |||||||||| || |||||||||||||||||||||||||||||||||||||||| ||||||
Db       1078 CCTATTCCAACAACCACCATCACAACTTCATATGTTGGTGTGACTACTTCCTACCTGACT 1137

Qy       1141 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1200
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 AAAACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1197

Qy       1201 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCA 1260
              ||||||||||||||||||||||||||||||||||||.||||||||||||||||||||||
Db       1198 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACTACAACTCGTACCAATCCA 1257

Qy       1261 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320
              |||||||| || || || || || || ||||||||||||||||||||||||| | ||||||
Db       1258 ACTGATTCTATAGATACTGTCGTTGTGCAAGTTCCACTGCCAAATCCAACTGTCACTACA 1317

Qy       1321 ACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGT 1380
              || || |||||||||||||| | |||||| || ||||| ||||||||||||||||||||||
Db       1318 ACCGAGTATTGGTCTCAGTCATATGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGT 1377

Qy       1381 ACCGATACTGTGATTATCAGAGAGCC 1406
              || ||  | || |||||||||| ||
Db       1378 ACTGACTCAGTAATTATCAGAGAACC 1403
```

```
RESULT 11
AF075293
LOCUS       AF075293               4332 bp    DNA     linear   PLN 03-AUG-2000
DEFINITION  Candida albicans strain 1161 agglutinin-like protein 6 (ALS6) gene,
            complete cds.
ACCESSION   AF075293
VERSION     AF075293.1  GI:5326751
KEYWORDS    .
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 4332)
  AUTHORS   Hoyer,L.L. and Hecht,J.E.
  TITLE     The ALS6 and ALS7 genes of Candida albicans
  JOURNAL   Yeast 16 (9), 847-855 (2000)
  MEDLINE   20321177
   PUBMED   10861907
REFERENCE   2  (bases 1 to 4332)
  AUTHORS   Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-1998) Veterinary Pathobiology, University of
            Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES             Location/Qualifiers
     source          1. .4332
                     /organism="Candida albicans"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="3"
     gene            1. .4332
                     /gene="ALS6"
     CDS             1. .4332
                     /gene="ALS6"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein 6"
                     /protein_id="AAD42033.1"
                     /db_xref="GI:5326752"
                     /translation="MKKVILLHLFFYCTIAMAKTISGVFTSFNSLTYTNTGNYPYGGP
                     GYPTWTAVLGWSLDGTLASPGDTFTLVMPCVFKFITTQTSVDLTANGVKYATCTFHAG
                     EDFTTFSSMSCVVNNGLSSNIRAFGTVRLPISFNVGGTGSSVNIQDSKCFTAGTNTVT
                     FTDGDHKISTTVNFPKTPQSSSSLVYFARVIPSLDKLSSLVVASQCTAGYASGVLGFS
                     ATKDDVTIDCSTIHVGITNGLNSWNMPVSSESFSYTKTCTPNSFIITYENVPAGYRPF
                     IDSYVKKSATATNGFNLNYTNIYNCMDGKKGNDPLIYFWTSYTNSDAGSNGAAVVVTT
                     RTVTDSTTAITTLPFDPTVDKTKTIEVIEPIPTTTITTSYVGISTSLSTKTATIGGTA
                     TVVVDVPYHTTTTITSIWTGSATTSSTYTNPTDSIDTVVVQVPSPNPTVTTTQFWSGS
                     VPTTETVTTGPQGTDSVIIKEPHNPTVTTTEFWSESFATTETVTNNPEGTDSVIIKEP
                     HNPTVTTTEFWSESFATTETVTNYPEGTDSVIVREPHNPTVTTTEFWSESFATTETVT
                     NYPEGTDSVIVREPHNPTVTTTEFWSESFATTETITNYPEGTDSVIVREPHNPTVTTT
                     EFWSESFVTTETITTGPLGTDSIVIHDPLEESSSSTAIESSDSNISSSAQESSSSVEQ
                     SLTSADETSSIVELSSRSDIPSSSIGLTSSESSTVSSYDSYSSSTSESSIASSYDSYS
                     SSSIESSTLSSSDRCSSSISDTTSFWDSSSSDLESTSITWSSSIDAQSSHLVQSVSNS
                     ISTSQELSSSSSEESSTFATDALVSSDASSILSSDTSSYYPSSTISSSDDFPHTIAGE
                     SDSLSISFITSTVEISSDSVSLTSDPASSFDSSSSLNSDSSSSPSSDQSDILTSSSFS
                     TLVVPSFSLSSSSSLSLTYPHYVNSTTYHASESESSSVASPSMASESANDDTYTLSES
                     TDTTSSIGTDSSTVTFCRRDNGDGCIVTGMPSSSIDSEQTSDVTTTSSFVASSTPTSA
                     EQSITDNPNIDSSQTSASSSTKLSVSVSDTVVNSISLSETSTLSSDDSTSSDTSISST
                     TNSDTGNVNAGSSHTSTASIKESSIQKTGVTLSSSYLSTKLSSTSDITTELITTELIT
                     TELTTTELTTIEDNEPNTFTSTPSSHSEIFSSDNSVLSKQVDGESTVEIPPVTDTTTV
                     SSVSVHSTEASTATLGENSFSKVASAPVNTETSLRSTSSSSNHATESSGTVKSEASAE
                     AIPSPPTSTDNRLSYSTEEAKGSTYPNSGSTNNLMTESQVAAPTDSTSVLTANPVVTS
                     TFDDKSSAAVNQPSKTKSIEESIGSLDSVNETNNGFIATLSQSEAPNSLIHSESISTT
                     MAKTTDASINGDSAASNSQPTTLIQQVATSSYNQPLITTYAGSSSATKHPSWLLKFIS
                     VALFFFL"
BASE COUNT     1324 a    966 c    752 g   1290 t
ORIGIN

  Query Match           23.9%;  Score 905.8;  DB 8;  Length 4332;
  Best Local Similarity   67.2%;  Pred. No. 9.3e-154;
  Matches 1298;  Conservative    0;  Mismatches  627;  Indels    6;  Gaps    1;

Qy         23 TATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATCACTGGTGTTTTTGATAGTTTTA 82
              | ||| | ||||   | || ||||  ||||| || ||  | || |||||    ||||| |
```

```
Db        26 TTTTCTTCTATTGCACGATAGCAATGGCAAAAACTATATCGGGAGTTTTCACGAGTTTCA 85

Qy        83 ATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTCAAAGGGCCAGGATACCCAACTT 142

Db        86 ACTCATTGACCTATACTAATACTGGTAACTACCCATATGGGGGTCCTGGTTATCCAACAT 145

Qy       143 GGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGTGCCAATCCAGGGGATACATTCA 202

Db       146 GGACTGCTGTTTTAGGTTGGAGCTTGGACGGAACACTAGCTAGTCCAGGTGATACATTTA 205

Qy       203 CATTGAATATGCCATGTGTGTTTAAATATACTACTTCACAAACATCTGTTGATTTAACTG 262

Db       206 CATTGGTCATGCCCTGCGTTTTCAAATTTATTACCACACAAACTTCAGTAGACTTAACTG 265

Qy       263 CCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCTGGTGAAGAATTCACAACTTTTT 322

Db       266 CTAATGGTGTCAAGTATGCAACATGTACTTTCCATGCAGGGGAAGACTTTACTACTTTTT 325

Qy       323 CTACATTAACATGTACTGTGAACGACGCTTTGAAATCATCCATTAAGGCATTTGGTACAG 382

Db       326 CAAGTATGAGTTGTGTAGTAAATAATGGGCTATCTTCAAATATCAGAGCGTTTGGTACCG 385

Qy       383 TTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGTTCATCAACTGATTTGGAAGATT 442

Db       386 TCAGGCTACCAATTTCATTCAATGTGGGTGGAACTGGTTCATCTGTCAACATTCAAGATT 445

Qy       443 CTAAATGTTTTACTGCTGGTACCAATACAGTCACATTTAATGATGGTGATAAAGATATCT 502

Db       446 CAAAGTGTTTCACTGCTGGAACGAACACTGTAACATTTACAGACGGCGATCACAAAATTT 505

Qy       503 CAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCAAGTGCATATTTGTATGCTTCCA 562

Db       506 CTACTACAGTCAATTTCCCTAAGACTCCACAATCATCTAGTAGCTTGGTTTATTTCGCAA 565

Qy       563 GAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTTGTGGCACCACAATGTGAAAATG 622

Db       566 GGGTTATTCCAAGTCTTGATAAATTATCTAGTCTTGTTGTTGCTTCTCAGTGTACTGCTG 625

Qy       623 GTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAACGGTGACGTTGCTATTGATTGCT 682

Db       626 GATATGCATCCGGTGTGCTCGGATTTTCAGCAACAAAAGATGATGTGACAATTGATTGTT 685

Qy       683 CAAATATTCATATTGGTATCACAAAAGGATTAAATGATTGGAATTATCCGGTTTCATCTG 742

Db       686 CTACTATACATGTGGGAATAACAAATGGTTTGAATAGTTGGAATATGCCAGTATCATCAG 745

Qy       743 AATCATTTAGTTACACTAAAACTTGTACATCTAATGGAATTCAGATTAAATATCAAAATG 802

Db       746 AATCATTTTCTTACACCAAAACTTGTACACCAAACAGTTTTATTATTACTTATGAAAATG 805

Qy       803 TACCTGCTGGTTATCGTCCATTTATTGATGCTT------ATATTTCTGCTACAGATGTTA 856

Db       806 TTCCTGCAGGTTATCGTCCATTTATTGATTCTTACGTGAAAAAATCAGCAACAGCAACGA 865

Qy       857 ACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGCAGTCGTCTGCAAAGTA 916

Db       866 ATGGATTTAATTTGAATTACACGAATATATACAATTGTATGGATGGCAAAAAGGGAAATG 925

Qy       917 AACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCCGGATCTAACGGTATTG 976

Db       926 ATCCTCTTATATACTTTTGGACATCATACACAAATAGTGATGCAGGATCCAATGGAGCTG 985

Qy       977 TCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCTGTCACTACTTTACCAT 1036

Db       986 CCGTAGTTGTTACTACGAGAACAGTCACTGATTCTACAACAGCAATTACCACATTACCGT 1045

Qy      1037 TCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAACCTATTCCAACCACTA 1096

Db      1046 TTGATCCCACAGTTGATAAAACCAAAACCATTGAAGTAATAGAACCCATCCCTACTACCA 1105

Qy      1097 CCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACTAAGACTGCACCAATTG 1156

Db      1106 CTATTACTACTTCATATGTTGGGATTTCTACTTCACTTTCTACGAAGACTGCAACTATTG 1165

Qy      1157 GTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACCACAACTGTTACCAGTG 1216

Db      1166 GAGGAACAGCAACTGTTGTTGTTGATGTTCCCTATCATACAACTACCACTATCACTAGTA 1225

Qy      1217 AATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCAACTGATTCAATTGACA 1276

Db      1226 TATGGACTGGATCAGCTACCACATCAAGTACTTATACAAATCCCACTGACTCGATTGATA 1285

Qy      1277 CAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACTACTGAATATTGGTCTC 1336

Db      1286 CAGTTGTTGTACAAGTTCCACTGCCAAATCCAACAGTTACAACTACTCAGTTTTGGTCAG 1345

Qy      1337 AGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTA 1396

Db      1346 GAAGTGTGCCCACAACCGAAACTGTGACCACTGGACCACAAGGAACGGATAGTGTGATCA 1405

Qy      1397 TCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTA 1456

Db      1406 TCAAGGAGCCACACAACCCTACTGTGACTACCACTGAGTTTTGGTCAGAATCATTTGCTA 1465

Qy      1457 CTACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCAC 1516

Db      1466 CTACTGAGACAGTCACCAACAATCCCGAAGGCACTGATAGTGTGATCATCAAGGAACCAC 1525

Qy      1517 CAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAG 1576

Db      1526 ACAATCCTACTGTTACCACCACCGAGTTTTGGTCAGAATCATTTGCTACTACTGAGACAG 1585

Qy      1577 TTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCTCCAAACCCAACTG 1636

Db      1586 TCACCAACTATCCAGAGGGAACTGACTCAGTCATTGTTAGAGAACCACACAATCCAACTG 1645

Qy      1637 TCACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACTACTGTGACTGCTCCTC 1696

Db      1646 TAACAACAACCGAGTTTTGGTCAGAATCATTTGCTACTACTGAGACAGTCACCAACTATC 1705

Qy      1697 CAGGAGGCACTGACTCAGTAATTATCAGAGAACCACCAAACCACACTGTCACTACTACTG 1756

Db      1706 CAGAGGGAACTGACTCAGTCATTGTTAGGGAACCACACAATCCTACTGTTACCACCACCG 1765

Qy      1757 AATACTGGTCACAATCATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTG 1816

Db      1766 AGTTTTGGTCAGAGTCATTTGCTACTACTGAGACCATCACCAACTATCCAGAGGGAACTG 1825

Qy      1817 ACACTGTTATCATTAGAGAGCCACCAAACCACACTGTCACTACTACTGAGTATTGGTCTC 1876

Db      1826 ACTCAGTCATTGTTAGGGAACCACACAATCCAACTGTAACAACAACCGAGTTTTGGTCAG 1885

Qy      1877 AATCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGCACTGATACTGTTATCA 1936

Db      1886 AATCATTTGTCACTACTGAAACGATAACTACAGGGCCACTTGGCACTGATAGTATCGTTA 1945

Qy      1937 TTAGGGAACCA 1947

Db      1946 TACATGATCCA 1956
```

```
RESULT 12
AF201684
LOCUS       AF201684               6897 bp    DNA     linear   PLN 09-AUG-2000
DEFINITION  Candida albicans agglutinin-like protein Als7p (ALS7) gene,
            complete cds.
ACCESSION   AF201684
VERSION     AF201684.1  GI:9754770
KEYWORDS    .
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 6897)
  AUTHORS   Hoyer,L.L. and Hecht,J.E.
  TITLE     The ALS6 and ALS7 genes of Candida albicans
```

```
  JOURNAL   Yeast 16 (9), 847-855 (2000)
  MEDLINE   20321177
   PUBMED   10861907
REFERENCE   2  (bases 1 to 6897)
  AUTHORS   Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1999) Veterinary Pathobiology, University of
            Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES             Location/Qualifiers
     source          1. .6897
                     /organism="Candida albicans"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="3"
     gene            <1. .>6897
                     /gene="ALS7"
     mRNA            <1. .>6897
                     /gene="ALS7"
                     /product="agglutinin-like protein Als7p"
     CDS             1. .6894
                     /gene="ALS7"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein Als7p"
                     /protein_id="AAF98068.1"
                     /db_xref="GI:9754771"
                     /translation="MKKLYLLYLLASFTTVISKEVTGVFNQFNSLIWSYTYRARYEEI
                     STLTANAQLEWALDGTIASPGDTFTLVMPCVYKFMTYETSVQLTANSIAYATCDFDAG
                     EDTKSFSSLKCTVTDELTEDTSVFGSVILPIAFNVGGSGSKSTITDSKCFSSGYNTVT
                     FFDGNNQLSTTANFLPRRELAFGLVVSQRLSMSLDTMTNFVMSTPCFMGYQSGKLGFT
                     SNDDDFEIDCSSIHVGITNEINDWSMPVSSVPFDHTIRCTSRALYIEFKTIPAGYRPF
                     VDAIVQIPTTEPFFVKYTNEFACVNGIYTSIPFTSFFSQPILYDEALAIGADLVRTTS
                     TVIGSITRTTTLPFISRLQKTKTILVLEPIPTTTVTTSHHGFDTWYYTKKATIGDTAT
                     VFIDVPQHTATTLTTYWQESSTATTTYFDDIDLVDTVIVKIPYPNPTIITTQFWSGKY
                     LTTETHKEPPLGTDSVIIKEPHNPTVTTTEFWSESFATTETITNYPEGTDSVIVREPH
                     NPTVTTTEFWSESFATTETITNGPEGTDSVIVREPHNPTVTTTEFWSESFATTETITN
                     GPEGTDSVIVREPHNPTVTTTEFWSESFATTETITNGPEGTDSVIVREPHNPTVTTTE
                     FWSESFATTETITNGPEGTDSVIIKEPHNPTVTTTKFWSESFATTETITNYPEGTDSV
                     IVREPHNPTVTTTEFWSESFATTETITNGPEGTDSVIVREPHNPTVTTTEFWSESFAT
                     TETITNYPEGTDSVIVREPHNPTVTTTEFWSESFATTETVTNYPEGTDSVIVREPHNP
                     TVTTTEFWSESFVTTETITTGPLGTDSIVIHDPLEESSSSTAIESSDSNISSSAQESS
                     SSVEQSFTSADETSSIVELSSRSDIPSSSIGLTSSESSTVSSYDSYSSSTSESSIASS
                     YDSYSSSSIESSTLSSSDRYSSSISDTTSFWDSSSSDLESTSITWSSSIDAQSSHLVQ
                     SVSNSISTSQEISSSSSEESSTSATDALVSSDASSILSSDTSSYYPSSTISPSDDFPH
                     TIAGESDSQSISFITSTVEISSDSVSLTSDPESSFDSSSRLNSDSSSSPSTDQRDILT
                     SSSFSTLIKSSGSRESSIGTILSEESSDSIPTTFSTRYWSPSGMSSRHYTNSTETSVS
                     DVVSSSVAGDETSESSVSVISESSESVTSESVASESVASESVASESVTAVSDI
                     SDLYTTSEEVSTSDSNSGMSSPIPSSEQRSSIPIMSSSDESSESRESSSGTILSEENS
                     DSIPTTFSTRYWSPSGMSSRHYTNSTETSVSDVVSSSVAGDETSESSVSVISESSESV
                     TSESVASESVASESVASESVTAVSDISDLYTTSEVVSTSDSNSGMSSPIPSSEQRSSI
                     PVMSSSDESSESRESSSGTILSEENSDSIPTTFSTRYLSPSGMSSRHYTNSTETSVSD
                     VVSSSVAGDETSESSVSVISESSESVTSESVASESVASESVASESVTAVSDISDLYTT
                     SEVVSTSDSKIVASTSVPSSEQRSSIPIMSSSDESSESRESSSGTILSEENSDSIPTT
                     FSTRYWSPSGMSSRHYTNSTETSVSDVVSSSVAGDETSESSVSVISESSESVTSESVA
                     SESVASESVTAVSDISDLYTTSEVVSTSDSKIVPSTSVPSSEQRSSIPIMSSSDESSE
                     SRESSSGTILSEENSDSIPTTFSTRYWSPSGMSSRHYTNSTETSVSDVVSSSVAGDET
                     SESSVSVTSESSESVTSESVASESVASESVASESVTAVSDISDLYTTSEVVSTSDSNS
                     GMSPPIPSSEQRSSIPVMSSSNESSESRESSSGTILSEENSDSIPTTFSTRYVSVSLT
                     VGELSALPSLPGKLSHLPSSLSETSIGMTKSANLSPQFFSTSVDSALSYWASGSSSAD
                     HQSSATCDVSESSVEGNLSAMAPGMSNSDDGLSEDTRSSSVAGKEEIELTSTNSVGEI
                     TLISYSSSSPTTHDHGRVSKSMGAAPLSSLFSVSVHTPLVTGLSGSDTFPSENSNRSR
                     SFKESTDNTISISRESLGNPYSSISSPSDYDVKSFTTSRELVSSESILPFSDVMDAND
                     MPTSGSNLHSMVFSISVLGEKFNANIEKHKNTNGHYSSMVFTYQSAGLEESDQRIAVT
                     NTKFDQNKIDTTIDSNTFVTSLPFATTSNDQIDQAVPIKIPASSTAGFVSDVLKPDYS
                     KSVQAESVQTDSTTYSEMMSSKRNKNSGFGTSSLILKPTITVVTKSIDTKVNTMKEGG
                     VSKQVSTTVTEQYDTSTYTPASLLVSDNSGSVSKYSLWMMAFYMLFGLF"
     misc_feature    1309. .2484
                     /gene="ALS7"
                     /note="Region: central tandem repeat domain"
     misc_feature    1309. .1425
                     /gene="ALS7"
                     /note="Region: ALS5-like tandem repeats"
     misc_feature    3292. .5511
                     /gene="ALS7"
                     /note="Region: VASES repeat region"
BASE COUNT     2020 a    1466 c    1411 g    2000 t
ORIGIN

  Query Match             23.1%;  Score 874.4;  DB 8;  Length 6897;
  Best Local Similarity   57.7%;  Pred. No. 3.9e-148;
  Matches 1609;  Conservative    0;  Mismatches 1166;  Indels   15;  Gaps    2;

Qy          7 CAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATCACTGGT 66
              | | |  ||  ||  || ||        ||  |||  |  |   | ||   | | ||||||
Db         10 CTATATCTTTTATATTTGTTGGCCCTGTTTACAACGGTCATATCCAAAGAAGTTACTGGT 69

Qy         67 GTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTCAAAGGG 126
              |||||  |    || |||||||| |  |||||  |  |    |    |   | | |||
Db         70 GTTTTCAACCAATTCAATTCATTGATATGGTCTTACACATACAGAGCTCGATACGAAGAA 129

Qy        127 CCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGTGCCAAT 186
                |     |  ||    ||||||  ||||  ||| | || ||||||||| | |||||
Db        130 ATATCTACTCTTACCGCTAATGCTCAATTGGAATGGGCTTTGGATGGTACTATTGCCAGT 189

Qy        187 CCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCACAAACA 246
              || || |||||||| |||||     ||||| ||||| | |||||  |  || |    ||||
Db        190 CCCGGTGATACATTTACATTAGTCATGCCCTGTGTATATAAATTCATGACGTACGAAACC 249

Qy        247 TCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCTGGTGAA 306
              || ||  | |||||||| |   | ||  |||||| || ||| | ||| || ||||||||
Db        250 TCAGTGCAATTAACTGCCAACTCTATTGCATATGCCACATGTGACTTTGATGCTGGTGAA 309

Qy        307 GAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCATCCATT 366
              ||    | || |||||| |  || |  |||.|| |||| || |  |||| |  |    |
Db        310 GACACTAAAAGTTTTTCAAGTTTGAAGTGTACGGTGACTGATGAGTTGACAGAAGATACC 369

Qy        367 AAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGTTCATCA 426
              |  |  ||||| |  |||| ||| || ||||| ||||||||||| ||  | || ||   |
Db        370 AGCGTTTTTGGAAGTGTTATTTTGCCTATTGCTTTCAATGTTGGAGGTTCCGGATCTAAA 429

Qy        427 ACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTTAATGAT 486
              ||    |  ||| || |||||||| |   |||  ||| || ||||| |||  |||
Db        430 TCTACGATAACAGACTCCAAATGTTTTTCAAGTGGGTACAACACTGTCACGTTTTTTGAC 489

Qy        487 GGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCAAGTGCA 546
              ||  | ||  | | || |   ||  | |||      ||  |     |  ||
Db        490 GGAAACAATCAACTTTCTACAACTGCAAATTTTCTTCCCCGAAGAGAACTAGCGTTTGGT 549

Qy        547 TATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTTGTGGCA 606
                |  |  |  ||| ||       ||| |||  | || | | || | | || | |
Db        550 CTAGTTGTTAGTCAAAGACTTTCCATGTCGCTCGATACAATGACTAATTTTGTTATGTCT 609

Qy        607 CCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAACGGTGAC 666
              |||  |||   |  |||||       ||||   | || ||  |   || | | | |||
Db        610 ACACCTTGTTTCATGGGTTATCAGCTGGGTAAGTTAGGTTTTACATCTAATGATGATGAT 669

Qy        667 GTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGATTGGAAT 726
              |||  |||||||| ||   ||| ||| ||||||| || || | | ||||||||||| |
Db        670 TTTGAAATTGATTGTTCTTCTATACATGTTGGTATAACTAATGAAATAAATGATTGGAGT 729

Qy        727 TATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGAATTCAG 786
                || || ||  |||| | ||  | | |||| ||  ||||||||   || | | |
Db        730 ATGCCAGTATCTTCTGTTCCCTTCGATCATACTATAAGATGTACATCACGTGCACTTTAC 789

Qy        787 ATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATTTCTGCT 846
              ||| | | | |||   | ||||| |||||||| || ||| | |||||   | ||
Db        790 ATTGAGTTTAAAACAATTCCTGCAGGTTATCGACCTTTTGTGGATGCGATTGTTCAAATA 849

Qy        847 ACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGCAGTCGT 906
              ||        | |  | |  | ||  |||||| ||||| | | | ||||      |
Db        850 CCAACGACAGAACCTTTTTTTGTAAAATATACAAATGAGTTTGCCTGTGTGAATGGCATA 909

Qy        907 CTGCAAAGTAAACCTTTCAC---TTTAAGATGGACTGGATACAAGAATAGTGATGCCGGA 963
                          | |||||||||  |||    |       |       ||  || ||   |
Db        910 TACACGTCCATACCTTTCACAAGTTTCTTTTCTCAGCCAATTTTATATGACGAGGCTTTA 969
```

```
Qy      964 TCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCTGTC 1023
Db      970 GCTATTGGTGCAGACCTAGTTCGTACCACATCCACAGTGATAGGTTCCATTACCAGAACT 1029

Qy     1024 ACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAACCT 1083
Db     1030 ACCACATTACCCTTCATTTCCCGACTCCAGAAAACCAAAACAATTCTAGTCTTAGAGCCC 1089

Qy     1084 ATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACTAAG 1143
Db     1090 ATACCCACCACTACGGTAACAACTTCACACCATGGCTTTGATACTTGGTATTATACTAAG 1149

Qy     1144 ACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACCACA 1203
Db     1150 AAAGCCACCATTGGTGACACAGCTACTGTTTTCATTGATGTTCCACAACATACAGCTACT 1209

Qy     1204 ACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCAACT 1263
Db     1210 ACTTTGACCACATATTGGCAAGAATCAAGTACAGCGACAACCACTTACTTCGATGACATA 1269

Qy     1264 GATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACTACT 1323
Db     1270 GACTTGGTCGATACTGTCATTGTGAAAATTCCATATCCCAATCCGACTATTATAACAACA 1329

Qy     1324 GAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGTACC 1383
Db     1330 CAATTTTGGTCAGGTAAATATTTAACTACTGAGACACACAAAGAACCACCTCTCGGTACT 1389

Qy     1384 GATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATATTGGTCA 1443
Db     1390 GATAGTGTGATCATCAAGGAACCACACAACCCTACTGTGACAACGACCGAGTTTTGGTCA 1449

Qy     1444 CAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATT 1503
Db     1450 GAATCATTTGCCACTACTGAGACCATCACCAACTATCCAGAAGGCACTGACTCAGTCATT 1509

Qy     1504 ATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCT 1563
Db     1510 GTTAGGGAACCACACAACCCTACTGTGACAACGACCGAGTTTTGGTCAGAATCATTTGCC 1569

Qy     1564 ACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCT 1623
Db     1570 ACTACTGAGACCATCACCAACGGTCCAGAAGGCACTGACTCAGTCATTGTTAGGGAACCA 1629

Qy     1624 CCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACTACT 1683
Db     1630 CATAATCCAACTGTGACAACAACCGAGTTTTGGTCAGAATCATTTGCCACTACTGAGACC 1689

Qy     1684 GTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAACCACCAAACCACACT 1743
Db     1690 ATCACCAACGGTCCAGAAGGCACTGACTCAGTCATTGTTAGGGAACCACACAATCCAACT 1749

Qy     1744 GTCACTACTACTGAATACTGGTCACAATCATATGCCACCACTACCACTGTAACTGCACCA 1803
Db     1750 GTGACAACAACCGAGTTTTGGTCAGAATCATTTGCCACTACTGAGACCATCACCAACGGT 1809

Qy     1804 CCAGGTGGTACTGACACTGTTATCATTAGAGAGCCACCAAACCACACTGTCACTACTACT 1863
Db     1810 CCAGAAGGCACTGACTCAGTCATTGTTAGGGAACCACATAATCCAACTGTGACAACAACC 1869

Qy     1864 GAGTATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGCACT 1923
Db     1870 GAGTTTTGGTCAGAATCATTTGCCACTACTGAGACCATCACCAACGGTCCAGAAGGCACT 1929

Qy     1924 GATACTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACTACTGAATACTGGTCT 1983
Db     1930 GATAGTGTGATCATCAAGGAACCACACAACCCTACTGTGACTACCACCAAGTTTTGGTCA 1989

Qy     1984 CAATCATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGTTCTT 2043
Db     1990 GAATCATTTGCCACTACTGAGACAATCACCAACTATCCAGAAGGCACTGACTCAGTCATT 2049

Qy     2044 ATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCT 2103

Db     2050 GTTAGGGAACCACATAATCCAACTGTGACAACAACCGAGTTTTGGTCAGAATCATTTGCC 2109

Qy     2104 ACAACCACCACTGTTACTGCACCACCTGGTGAAACCGATACCGTTCTTATCAGAGAGCCA 2163
Db     2110 ACTACTGAGACCATCACCAACGGTCCAGAAGGCACTGACTCAGTCATTGTTAGGGAACCA 2169

Qy     2164 CCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACCACT 2223
Db     2170 CACAATCCAACTGTGACAACAACCGAGTTTTGGTCAGAATCATTTGCTACTACTGAGACC 2229

Qy     2224 GTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGAGCCACCAAATCCAACA 2283
Db     2230 ATCACCAACTATCCAGAGGGAACTGACTCAGTCATTGTTAGAGAACCACACAATCCAACT 2289

Qy     2284 GTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACAGTTACTGCTCCT 2343
Db     2290 GTAACAACAACCGAGTTTTGGTCAGAGTCATTTGCTACTACTGAGACAGTCACCAACTAT 2349

Qy     2344 CCAGGTGGTACTGACACTGTGATTAT----------CTATGAAAGCATGTCAAGTTCA 2391
Db     2350 CCAGAAGGCACTGACTCAGTCATTGTTAGGGAACCACACAATCCAACTGTGACAACAACC 2409

Qy     2392 AAGATTTCTACATCCTCCAATGATATAACCAGTATCATTCCATCATTTTCCCGTCCTCAT 2451
Db     2410 GAGTTTTGGTCAGAATCATTTGTCACTACTGAAACGATAACTACAGGGCCACTTGGCACT 2469

Qy     2452 TATGTCAACAGCACAACCTCCGATTTGTCAACATTTGAATCTTCATCCATGAATACTCCT 2511
Db     2470 GATAGTATCGTTATACATGATCCATTGGAAGAACTGTCTTCTAGTACTGCTATTGAGTCA 2529

Qy     2512 ACTTCTATCAGTAGTGATGGTATGTTGTTGTCTTCTACAACTTTGGTTACTGAATCAGAA 2571
Db     2530 AGTGATTCTAATATTTCAAGCTCAGCTCAAGAATCATCCAGTCTGGTTGAACAGTCATTT 2589

Qy     2572 ACAACTACAGAACTGATTTGCAGTGATGGTAAAGAGTGTTCTAGATTGTCCAGTTCTTCT 2631
Db     2590 ACTTCTGCTGACGAGACTTCAAGTATAGTTGAATTGTCATCAAGATCAGACATTCCATCA 2649

Qy     2632 GGTATTGTCACAAATCCAGATAGCAATGAATCCTCAATCGTAACTAGTACTGTTCCTACT 2691
Db     2650 AGCTCAATTGGGTTAACATCTAGTGAGTCGTCTACTGTCTCAAGTTATGATAGCTACTCT 2709

Qy     2692 GCAAGTACAATGTCTGATTCACTTTCTTCAACTGATGGTATTAGTGCTACATCTTCTGAT 2751
Db     2710 TCAAGTACTAGCGAATCATCTATTGCTTCAAGTTATGATAGCTATTCGTCAAGTAGTATT 2769

Qy     2752 AATGTTTCAAAATCAGGAGTATCAGTTACA 2781
Db     2770 GAGTCGTCTACTTTATCTAGTTCCGATAGA 2799
```

```
RESULT 13
CANALS4S1
LOCUS       CANALS4S1               1407 bp    DNA     linear   PLN 16-OCT-1998
DEFINITION  Candida albicans agglutinin-like protein (ALS4) gene, 5' partial
            cds.
ACCESSION   AF024584
VERSION     AF024584.1  GI:3598677
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 1407)
  AUTHORS   Hoyer,L.L., Payne,T.L. and Hecht,J.E.
  TITLE     Identification of Candida albicans ALS2 and ALS4 and localization
            of als proteins to the fungal cell surface
  JOURNAL   J. Bacteriol. 180 (20), 5334-5343 (1998)
  MEDLINE   98440424
   PUBMED   9765564
REFERENCE   2  (bases 1 to 1407)
```

```
  AUTHORS   Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1997) Veterinary Pathobiology, University of
            Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES             Location/Qualifiers
     source          1. .1407
                     /organism="Candida albicans"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="6"
                     /map="6C"
     gene            1. .>1407
                     /gene="ALS4"
                     /note="agglutinin-like sequence"
                     /allele="ALS4-1"
     CDS             1. .>1407
                     /gene="ALS4"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAC64239.1"
                     /db_xref="GI:3598680"
                     /translation="MLLQFLLLSLCVSVATAKVITGVFNSFNSLTWANAASYPYRGPA
                     TPTWTAVIGWSLDGATASAGDTFTLDMPCVFKFITDQTSIDLVADGRTYATCNLNSAE
                     EFTTFSSVSCTVTTTMTADTKAIGTVTLPFSFSVGGSGSDVDLANSQCFTAGINTVTF
                     NDGDTSISTTVDFEKSTVASSDRILLSRILPSLSQAVNLFLPQECANGYTSGTMGFST
                     AGTGATIDCSTVHVGISNGLNDWNYPISSESFSYTKTCTSTSVLVTFQNVPAGYRPFV
                     DAYISATRVSSYTMQYTNIYACVGAASVDDSFTHTWRGYSNSQAGSNGITIVVTTRTV
                     TDSTTAVTTLPFNSDTDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPSPNPTVTTTEYWSQSYAT
                     TTTVTAPPGGTDSVIIREP"
     repeat_region   1300. .1407
                     /note="first of multiple copies of 108bp imperfect repeat"
                     /rpt_type=tandem
BASE COUNT      391 a    294 c    253 g    469 t
ORIGIN

  Query Match             22.3%;  Score 842.8;  DB 8;  Length 1407;
  Best Local Similarity   75.0%;  Pred. No. 2.8e-142;
  Matches 1054;  Conservative    0;  Mismatches  352;  Indels    0;  Gaps    0;

Qy          1 ATGCTTCAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATC 60
              ||||||  |||||||   ||| ||  ||| | | | ||| |||| |  |||||    ||
Db          1 ATGCTTTTACAATTTTTGTTGCTAAGCCTCTGTGTATCAGTTGCTACGGCAAAAGTTATT 60

Qy         61 ACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTC 120
              ||  ||||||||  |||||||||||||| |||||||||| ||||||||||  |||| | |
Db         61 ACAGGTGTTTTCAATAGTTTTAATTCGTTAACTTGGGCCAATGCTGCTTCTTATCCATAT 120

Qy        121 AAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGT 180
              | ||| ||||     || |||||||  |||||  | || ||||| ||||||||  | | |
Db        121 AGAGGTCCAGCTACTCCTACTTGGACCGCTGTAATAGGATGGTCTTTAGATGGAGCTACT 180

Qy        181 GCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCA 240
              || | | || || |||||||| ||  | ||||| ||||| || |||| || ||||
Db        181 GCTAGTGCTGGTGACACATTCACGTTAGACATGCCTTGTGTTTTCAAATTTATTACTGAT 240

Qy        241 CAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 300
              |||||  ||  |||||||  ||| ||||||  ||  |||||||||||| | ||  |||||
Db        241 CAAACGTCAATTGATTTAGTTGCTGATGGTCGTACTTATGCTACTTGTAATTTGAATTCT 300

Qy        301 GGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 360
              |  |||||  ||  ||  ||||||||||   |  ||||||||||||      |  |||   |
Db        301 GCCGAAGAGTTTACTACTTTTTCTAGTGTGTCATGTACTGTGACTACTACAATGACTGCT 360

Qy        361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 420
                ||  ||  ||  | ||  ||  ||  ||  |||||  |  ||||||  |||  ||  |||  |||||
Db        361 GACACCAAAGCCATAGGAACTGTAACATTACCTTTCTCATTCAGTGTGGGGGGATCAGGT 420

Qy        421 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTT 480
              |||     ||||||||  | |||||  ||||||||||||  ||  | |||||||||  ||  |||
Db        421 TCAGATGTTGATTTGGCAAATTCTCAATGTTTTACTGCAGGAATCAATACAGTTACTTTT 480

Qy        481 AATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCA 540
              ||||||||||| |      || || |     ||||| |||||||| |||||||| |   |
Db        481 AATGATGGTGACACTAGCATTTCCACAACAGTTGATTTTGAAAAATCAACCGTGGCCTCC 540

Qy        541 AGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTT 600
              || |    | | |      || ||| || | || |||||      | |    || |||||||
Db        541 AGCGATCGTATCTTGTTGTCAAGAATTTTACCCAGTCTTTCACAAGCAGTAAATCTTTTT 600

Qy        601 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC 660
              |  | | | |||||| |||||||| || ||||||||||||| || || | |  |
Db        601 CTTCCCCAAGAATGTGCAAATGGTTATACTTCTGGTACAATGGGATTTTCGACTGCTGGT 660

Qy        661 GGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGAT 720
                ||  | | |||| ||||| || |  ||||| | || ||  |||| || || ||||||
Db        661 ACTGGTGCTACTATAGATTGTTCCACAGTTCATGTCGGGATATCAAATGGGTTGAATGAT 720

Qy        721 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 780
              ||||||||||  |||||   ||||| ||| ||||||  || || |||||||| |   |
Db        721 TGGAATTATCCAATTTCACTGGAATCTTTTTCTTACACAAAGACCTGTACATCAACAAGT 780

Qy        781 ATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATT 840
              ||    | |  | |||||||| ||||| || |||||||||||| ||||||||||||||
Db        781 GTTTTAGTAACTTTTCAAAATGTTCCTGCCGGATATCGTCCATTTGTTGATGCTTATATT 840

Qy        841 TCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGC 900
              ||||| |||   || | |  ||||||  |   ||| || |||   ||| |||||| ||||
Db        841 TCTGCAACACGAGTCAGCTCATATACCATGCAATACACTAATATATATGCTTGTGTTGGC 900

Qy        901 AGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCC 960
                    |       | |  | || |||   |  |||   ||||| |  |||||| | ||
Db        901 GCGGCTTCTGTTGATGACTCATTTACTCATACTTGGCGGGGATATAGTAATAGTCAAGCT 960

Qy        961 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1020
              || ||||| ||||||  ||||||  |  ||||||||||||||||||||||||||||||||
Db        961 GGTTCTAATGGTATTACCATTGTGGTAACAACTAGAACAGTTACAGACAGTACCACTGCT 1020

Qy       1021 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1080
              || |||||||||||||||||  |   |  ||| ||||||||||||||||||||||||| |||
Db       1021 GTGACTACTTTACCATTCAATTCCGATACTGACAAAACCAAAACAATCGAAATTTTACAA 1080

Qy       1081 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1140
              |||||||||| |||||||| ||||||||||||||||||||||| |||||||| ||||||
Db       1081 CCTATTCCAACAACTACCATTACAACTTCATATGTTGGTGTGACAACTTCCTACCTGACT 1140

Qy       1141 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1200
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAAACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACT 1200

Qy       1201 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCA 1260
              |||||||||||||||||||||||||||||||||||||||| ||||||||  ||||||||||||||||
Db       1201 ACAACTGTTACCAGTGAATGGACAGGAACAATTACTACCACTACAACTCGTACCAATCCA 1260

Qy       1261 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320
              |||||||| || || || || || || ||||||||||||||||||||||| | |||
Db       1261 ACTGATTCTATAGATACTGTCGTTGTTCAAGTTCCACTGCCAAATCCAACTGTCACTACA 1320

Qy       1321 ACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGT 1380
              || || |||||||||||||| | |||||| || ||||| ||||||||||||||||||||||
Db       1321 ACCGAGTATTGGTCTCAGTCATATGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGT 1380

Qy       1381 ACCGATACTGTGATTATCAGAGAGCC 1406
              || ||  | || |||||||||||| ||
Db       1381 ACTGACTCAGTAATTATCAGAGAACC 1406


RESULT 14
CANALS42S1
LOCUS       CANALS42S1              1407 bp    DNA     linear   PLN 16-OCT-1998
DEFINITION  Candida albicans agglutinin-like protein (ALS4) gene, 5' partial
            cds.
ACCESSION   AF024586
VERSION     AF024586.1  GI:3598682
KEYWORDS    .
```

```
SEGMENT     1 of 2
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 1407)
  AUTHORS   Hoyer,L.L., Payne,T.L. and Hecht,J.E.
  TITLE     Identification of Candida albicans ALS2 and ALS4 and localization
            of als proteins to the fungal cell surface
  JOURNAL   J. Bacteriol. 180 (20), 5334-5343 (1998)
  MEDLINE   98440424
   PUBMED   9765564
REFERENCE   2  (bases 1 to 1407)
  AUTHORS   Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1997) Veterinary Pathobiology, University of
            Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES             Location/Qualifiers
     source          1. .1407
                     /organism="Candida albicans"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="6"
                     /map="6C"
     gene            1. .>1407
                     /gene="ALS4"
                     /note="agglutinin-like sequence"
                     /allele="ALS4-2"
     CDS             1. .>1407
                     /gene="ALS4"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAC64241.1"
                     /db_xref="GI:3598685"
                     /translation="MLLQFLLLSLCVSVATAKVITGVFNSFNSLTWANAASYPYRGPA
                     TPTWTAVIGWSLDGATASAGDTFTLDMPCVFKFITDQTSIDLVADGRTYATCNLNSAE
                     EFTTFSSVSCTVTTTMTADTKAMGTVTLPFSFSVGGSGSDVDLANSQCFTAGINTVTF
                     NDGDTSISTTVDFEKSTVASSDRILLSRILPSLSQAVSLFLPQECANGYTSGTMGFST
                     AGTGATIDCSTVHVGISNGLNDWNYPISSESFSYTKTCTSTSVLVTYQNVPAGYRPFV
                     DAYVSATRVSSYTMQYTNIYACVGAASVDDSFTHTWRGYSNSQAGSNGITIVVTTRTV
                     TDSTTAVTTLPFNSDTDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPSPNPTVTTTEYWSQSYAT
                     TTTVTAPPGGTDSVIIREP"
     repeat_region   1300. .1407
                     /note="first of multiple copies of 108bp imperfect repeat"
                     /rpt_type=tandem
BASE COUNT      389 a    297 c    257 g    464 t
ORIGIN

  Query Match           22.1%;  Score 836.4;  DB 8;  Length 1407;
  Best Local Similarity   74.7%;  Pred. No. 4e-141;
  Matches 1050;  Conservative    0;  Mismatches  356;  Indels    0;  Gaps    0;

Qy          1 ATGCTTCAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATC 60
              ||||||  |||||||   ||| ||  ||| | | | ||| |||| |  |||||    ||
Db          1 ATGCTTTTACAATTTTTGTTGCTAAGCCTCTGTGTATCAGTTGCTACGGCAAAAGTTATT 60

Qy         61 ACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTC 120
              || ||||||||  |||||||||||||| |||||||| ||||||||||  |||| | |
Db         61 ACAGGTGTTTTCAATAGTTTTAATTCGTTAACTTGGGCCAATGCTGCTTCTTATCCATAT 120

Qy        121 AAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGT 180
              | ||| ||||     || |||||||  |||||  | || ||||| ||||||||  | | |
Db        121 AGAGGTCCAGCTACTCCTACTTGGACCGCTGTAATAGGATGGTCTTTAGATGGAGCTACT 180

Qy        181 GCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCA 240
              || | | | || || |||||||| ||  | ||||| ||||| || |||| || ||||
Db        181 GCTAGTGCTGGTGACACATTCACGTTAGACATGCCTTGTGTTTTCAAATTTATTACTGAT 240

Qy        241 CAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 300
              ||||| ||  ||||||||  ||| ||||||  ||  ||||||||||||| | ||  | |||
Db        241 CAAACGTCAATTGATTTAGTTGCTGATGGTCGTACTTATGCTACTTGTAATTTGAACTCT 300

Qy        301 GGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 360
              |  ||||| || || ||||||||||   |  ||||||||||||      |  |||   |
Db        301 GCCGAAGAGTTTACTACTTTTTCTAGTGTGTCATGTACTGTGACTACTACAATGACTGCT 360

Qy        361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 420
              ||  || ||  | || || || || ||||| |  |||||| ||| || ||| |||||
Db        361 GACACCAAAGCCATGGGAACTGTAACATTACCTTTCTCATTCAGTGTGGGGGGATCAGGT 420

Qy        421 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTT 480
              |||     |||||||| | ||||| ||||||||||||| || | |||||||| || |||
Db        421 TCAGATGTTGATTTGGCAAATTCTCAATGTTTTACTGCAGGAATCAATACAGTTACTTTT 480

Qy        481 AATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCA 540
              |||||||||| |     || || |     ||||| |||||||| |||||||| |   |
Db        481 AATGATGGTGACACTAGCATTTCCACAACAGTTGATTTTGAAAAATCAACCGTGGCCTCC 540

Qy        541 AGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTT 600
              || |    | | |    || ||| || | || |||||     | |    || |||||||
Db        541 AGCGATCGTATCTTGTTGTCAAGAATTTTACCCAGTCTTTCACAAGCAGTAAGTCTTTTT 600

Qy        601 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC 660
              |  | | | |||||| |||||||| || |||||||||||||| || || | |  |
Db        601 CTTCCCCAAGAATGTGCAAATGGTTATACTTCTGGTACAATGGGATTTTCAACTGCTGGT 660

Qy        661 GGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGAT 720
              ||  | | |||| ||||| || |  |||| | || ||  |||| || || ||||||
Db        661 ACTGGTGCTACTATAGATTGTTCCACAGTTCATGTCGGGATATCAAATGGGTTGAATGAT 720

Qy        721 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 780
              ||||||||||| |||||  ||||| ||| |||||| || || |||||||| |   |
Db        721 TGGAATTATCCAATTTCACTGGAATCTTTTTCTTACACAAAGACCTGTACATCAACAAGT 780

Qy        781 ATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATT 840
              ||    | |  ||||| ||||| ||||| || |||||||||||| | |||||||| ||
Db        781 GTTTTAGTAACTTATCAGAATGTTCCTGCCGGATATCGTCCATTTGTCGATGCTTATGTT 840

Qy        841 TCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGC 900
              ||||| |||   || | |  |||||| |   ||| || |||   ||| |||||| ||||
Db        841 TCTGCAACACGAGTCAGCTCATATACCATGCAATACACTAATATATATGCTTGTGTTGGC 900

Qy        901 AGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCC 960
              |          | |  | || |||   | |||   ||||| |  |||||| | ||
Db        901 GCGGCTTCTGTTGATGACTCATTTACTCATACTTGGCGGGGATATAGTAATAGTCAAGCT 960

Qy        961 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1020
              || ||||| ||||||  |||||| |  ||||||||||||||||||||||||||||||||
Db        961 GGTTCTAATGGTATTACCATTGTGGTAACAACTAGAACAGTTACAGACAGTACCACTGCT 1020

Qy       1021 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1080
              || |||||||||||||||| |   |  ||| |||||||||||||||||||||||| |||
Db       1021 GTGACTACTTTACCATTCAATTCCGATACTGACAAAACCAAAACAATCGAAATTTTACAA 1080

Qy       1081 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1140
              |||||||||| |||||||| |||||||||||||||||||| ||||||| ||||||
Db       1081 CCTATTCCAACAACTACCATTACAACTTCATATGTTGGTGTGACAACTTCCTACCTGACT 1140

Qy       1141 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1200
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AAAACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACT 1200

Qy       1201 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCA 1260
              |||||||||||||||||||||||||||||||| ||||||| ||||||||||||||||||
Db       1201 ACAACTGTTACCAGTGAATGGACAGGAACAATTACTACCACTACAACTCGTACCAATCCA 1260

Qy       1261 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320
              |||||||| || || || || || || ||||||||||||||||||||||| | |||
Db       1261 ACTGATTCTATAGATACTGTCGTTGTTCAAGTTCCACTGCCAAATCCAACTGTCACTACA 1320

Qy       1321 ACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGT 1380
              || || |||||||||||||| | |||||| || ||||| |||||||||||||| ||
Db       1321 ACCGAGTATTGGTCTCAGTCATATGCTACTACTACTGTTACTGCTCCTCCAGGAGGC 1380
```

```
Qy       1381 ACCGATACTGTGATTATCAGAGAGCC 1406
              || ||  | ||||||||||||||| ||
Db       1381 ACTGACTCAGTGATTATCAGAGAACC 1406


RESULT 15
AF229989S1
LOCUS       AF229989S1              1404 bp    DNA     linear   PLN 02-FEB-2001
DEFINITION  Candida albicans agglutinin-like protein (ALS9) gene, partial cds.
ACCESSION   AF229989
VERSION     AF229989.1  GI:12656144
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 1404)
  AUTHORS   Hoyer,L.L., Hecht,J.E. and Mirus,K.A.
  TITLE     The ALS9 gene of Candida albicans
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1404)
  AUTHORS   Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2000) Veterinary Pathobiology, University of
            Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES             Location/Qualifiers
     source          1. .1404
                     /organism="Candida albicans"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="6"
                     /map="6O"
     CDS             1. .>1404
                     /gene="ALS9"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAK00764.1"
                     /db_xref="GI:12656146"
                     /translation="MLPQFILLFISLTVSTAKTITGVFNSFDSLTWTRSVEYAYKGPE
                     TPTWNAVLGWSLNSTTADPGDTFTLILPCVFKFITTQTSVDLTADGVSYATCDFNAGE
                     EFTTFSSLSCTVNSVSVSYARVSGTVKLPITFNVGGTGSSVDLADSKCFTAGKNTVTF
                     MDGDTKISTTVDFDASPVSPSGYITSSRIIPSLNKLSSLFVVPQCENGYTSGIMGFVA
                     SNGATIDCSNVNIGISKGLNDWNFPVSSESFSYTKTCTSTSITVEFQNVPAGYRPFVD
                     AYISAENIDKYTLTYANEYTCENGNTVVDPFTLTWWGYKNSEADSDGDVIVVTTRTVT
                     DSTTAVTTLPFNPSVDKTETIEILQPIPTTTIITSYIGISTSYETLTGTIGGTATVIV
                     DTPYHITATVTNFWTGSITTTTTYTNPTGSIDTVIVQIPSPDPTTTITEFWSESFAST
                     TTITNPPDGTNSVIIKEP"
     repeat_region   1297. .>1404
                     /note="first of a series of 108 bp repeat units"
                     /rpt_type=tandem
BASE COUNT      413 a    286 c    253 g    452 t
ORIGIN

  Query Match           21.3%;  Score 805.4;  DB 8;  Length 1404;
  Best Local Similarity   73.8%;  Pred. No. 1.6e-135;
  Matches 1038;  Conservative    0;  Mismatches  366;  Indels    3;  Gaps    1;

Qy          1 ATGCTTCAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATC 60
              ||||||| |||||| | |||||||||| ||| |||| || |  | | |||||| || ||
Db          1 ATGCTTCCACAATTCATATTGTTATTCATATCTTTGACAGTGTCGACTGCAAAAACTATT 60

Qy         61 ACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTC 120
              |||||||||| |||||||| | ||||| || ||| | |   | | | | ||||||| |
Db         61 ACTGGTGTTTTCAATAGTTTTGACTCATTGACATGGACTAGATCCGTTGAATATGCTTAC 120

Qy        121 AAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGT 180
              |||||||||     |||||||||||||| ||||| || ||||||||| || |||||| |
Db        121 AAAGGGCCAGAGACTCCAACTTGGAATGCAGTTTTAGGGTGGTCCTTAAATAGTACCACT 180

Qy        181 GCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCA 240
              ||  | ||||| || |||||||| |||| | |||| ||||| ||||||| ||  ||| |

Db        181 GCTGACCCAGGAGACACATTCACCTTGATTTTGCCTTGTGTATTTAAATTTATAACTACC 240

Qy        241 CAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 300
              ||||||||||||||||| ||||| |||||||||  ||||| |||||| | ||| || ||
Db        241 CAAACATCTGTTGATTTGACTGCTGATGGTGTTAGCTATGCCACTTGTGACTTTAATGCT 300

Qy        301 GGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 360
              |||||||||| || || |||||| | ||| ||||||||||||||    | || |  ||||
Db        301 GGTGAAGAATTTACGACATTTTCTTCCTTATCATGTACTGTGAACAGTGTTTCGGTATCA 360

Qy        361 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 420
              |    || ||  | |||||| || |  || || ||| |||||||||| ||||||||||
Db        361 TATGCTAGGGTTTCTGGTACGGTCAAATTGCCCATTACATTCAATGTAGGTGGAACAGGT 420

Qy        421 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTT 480
              || |||  |||||||| |||||| |||||||||||||| || |  || || || || ||
Db        421 TCTTCAGTTGATTTGGCAGATTCCAAATGTTTTACTGCCGGAAAAAACACTGTGACTTTC 480

Qy        481 AATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCA 540
              |  ||||| |||| | | || || |    |||||| |||||   ||| | ||    ||
Db        481 ATGGATGGCGATACAAAGATTTCTACCACTGTTGATTTTGACGCGTCTCCAGTATCACCC 540

Qy        541 AGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTT 600
              ||||  ||| |       ||  || |||| || |||||||||||  |  ||| ||||||
Db        541 AGTGGTTATATTACAAGCTCACGAATTATTCCTAGTCTCAATAAACTTTCAAGTCTTTTC 600

Qy        601 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC 660
              ||||  |||||||||| |||||||||||||||||||| |||||| ||      |||||||
Db        601 GTGGTGCCACAATGTGAGAATGGTTACACATCTGGTATAATGGGATTTGTAGCTAGTAAC 660

Qy        661 GGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGAT 720
              |||   | | |||||||||||||||  |||| || ||  ||||||| |||||||||
Db        661 GGT---GCTACTATTGATTGCTCAAATGTCAATATAGGAATATCAAAAGGTTTAAATGAT 717

Qy        721 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 780
              ||||||| ||| ||    || |||||||||  |||||| |||||||||| || |   |
Db        718 TGGAATTTTCCAGTAAGTTCAGAATCATTTTCTTACACGAAAACTTGTACGTCAACCAGT 777

Qy        781 ATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATT 840
              |||    || ||| |||||||| |||||||| ||||| || ||| ||||||| ||||||
Db        778 ATTACAGTTGAATTTCAAAATGTTCCTGCTGGGTATCGCCCTTTTGTTGATGCATATATT 837

Qy        841 TCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGC 900
              |||||   | || || |  ||||||| ||  | ||  | ||||| ||||||||||
Db        838 TCTGCAGAAAATATTGATAAATATACCTTGACGTACGCAAATGAGTATACTTGTGAAAAT 897

Qy        901 AGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCC 960
              |   |         | | || || ||||||| ||||   || || || ||   ||| ||
Db        898 GGCAATACTGTGGTTGATCCATTTACTTTAACATGGTGGGGGTATAAAAACTCTGAAGCA 957

Qy        961 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 1020
              |  ||| ||||   ||| || || | |||||| ||||| || |||||||||| || |||
Db        958 GACTCTGACGGGGATGTGATCGTAGTTACAACCAGAACTGTCACAGACAGTACAACAGCT 1017

Qy       1021 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1080
              || |||||||||| |||||||||||||| |||||||| ||||||| ||||||||||||
Db       1018 GTGACTACTTTACCTTTCAATCCAAGTGTCGATAAAACCGAAACAATTGAAATTTTGCAA 1077

Qy       1081 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1140
              || ||||| || || || || |||||||||||| ||||| |  | |||||||   ||
Db       1078 CCCATTCCCACGACCACAATTACAACTTCATATATTGGTATTTCCACTTCCTATGAAACA 1137

Qy       1141 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACC 1200
                || | | |||||||||  || || || || ||||| |||   || |||||||  ||
Db       1138 TTAACCGGAACAATTGGTGGTACTGCGACAGTCATTGTCGATACACCTTATCATATCACT 1197

Qy       1201 ACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCA 1260
              | |||||||| | |  ||||| ||  |||| || || ||||| |||  ||| |||||
Db       1198 GCCACTGTTACAAATTTCTGGACTGGGTCAATTACAACTACCACTACTTATACTAATCCC 1257

Qy       1261 ACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACT 1320
              |||| ||| || ||||| ||  | || ||| |||||||||  |||||||   | ||
Db       1258 ACTGGTTCCATAGACACTGTTATTGTGCAAATTCCACTGCCTGATCCAACTACAACTATA 1317
```

```
Qy      1321 ACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGT 1380
             ||||||| ||||||| | || |||||||  || || ||  | ||    || || |  |||
Db      1318 ACTGAATTTTGGTCTGAATCATTTGCTAGTACTACCACCATCACCAACCCACCTGACGGT 1377

Qy      1381 ACCGATACTGTGATTATCAGAGAGCCA 1407
             ||  ||| |||||| |||| ||| |||
Db      1378 ACTAATAGTGTGATCATCAAAGAACCA 1404
```

Search completed: June 20, 2003, 05:17:40
Job time : 6411 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 01:57:43 ; Search time 530 Seconds
(without alignments)
16086.916 Million cell updates/sec

Title: US-09-715-876-7
Perfect score: 3786
Sequence: 1 atgcttcaacaatttacatt.....tgtccttgttcatttagtga 3786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	359.2	9.5	424	17	AAT29067	Probe for Candida
2	108	2.9	108	17	AAT32337	Probe for Candida
3	108	2.9	108	17	AAT32330	Probe for Candida
4	108	2.9	108	17	AAT32331	Probe for Candida
5	108	2.9	108	17	AAT32332	Probe for Candida
6	108	2.9	108	17	AAT32334	Probe for Candida
7	108	2.9	108	17	AAT32336	Probe for Candida
8	108	2.9	108	17	AAT29063	Probe for Candida
9	106.4	2.8	108	17	AAT32333	Probe for Candida
10	106.4	2.8	108	17	AAT32335	Probe for Candida
11	99	2.6	99	17	AAT32338	Probe for Candida
c 12	94.2	2.5	8201	21	AAA88864	Human dentin sialo
c 13	94.2	2.5	8201	24	ABQ73537	Human dentin sialo
14	87	2.3	5511	21	AAA61847	Cryptosporidium pa
15	87	2.3	5511	24	ABT04776	C parvum GP900 gen
16	87	2.3	7334	21	AAA61846	Cryptosporidium pa
17	87	2.3	7334	24	ABT04775	C parvum GP900 gen
18	86.2	2.3	5163	19	AAV20700	Cryptosporidium pa
19	86.2	2.3	5163	21	AAA61849	ORF encoding a por
20	86.2	2.3	5163	24	ABT04778	C parvum GP900 gen
21	86.2	2.3	5318	19	AAV20701	Cryptosporidium pa
22	86.2	2.3	5318	21	AAA61848	DNA encoding a por
23	86.2	2.3	5318	24	ABT04777	C parvum GP900 gen
24	85.2	2.3	108	17	AAT29064	Probe for Candida
c 25	82.4	2.2	100	15	AAQ62588	Candida albicans-s
c 26	82.4	2.2	100	15	AAQ62590	Candida albicans-s
c 27	81.4	2.2	7758	24	ABL33102	Human immune syste
c 28	76	2.0	100	15	AAQ62589	Candida albicans-s
29	74.4	2.0	267156	24	ABL68560	Kidney cancer rela
c 30	68.4	1.8	40875	18	AAT80043	Insert from cosmid
c 31	68	1.8	100	15	AAQ62592	Candida albicans-s
c 32	68	1.8	7758	24	ABL33103	Human immune syste
c 33	67.6	1.8	2336	23	ABL25662	Drosophila melanog
c 34	64.8	1.7	2790	7	AAN60473	Sequence encoding
35	62.6	1.7	1368	10	AAN91235	DNA sequence of pr
c 36	62.2	1.6	2014	24	ABA90791	Bacillus anthracis
37	58	1.5	3168	20	AAX36552	C. albicans Rbt1 c
c 38	58	1.5	3983	23	ABL15838	Drosophila melanog
c 39	57.4	1.5	2535	23	ABL25476	Drosophila melanog
c 40	57.2	1.5	100	15	AAQ62593	Candida albicans-s
c 41	57.2	1.5	3399	17	AAT05868	Chicken leucocytoz
42	57.2	1.5	32392	24	ABL56203	AmEPV genome fragm
c 43	57.2	1.5	50000	24	ABL55643	AmEPV genome fragm
c 44	57	1.5	100	15	AAQ62591	Candida albicans-s
c 45	56.6	1.5	14987	24	ABL32630	Human immune syste

ALIGNMENTS

RESULT 1
AAT29067
ID AAT29067 standard; DNA; 424 BP.
XX
AC AAT29067;
XX
DT 28-NOV-1996 (first entry)
XX
DE Probe for Candida albicans and Candida stellatoides.
XX
KW Probe; primer; detection; identification; Candida albicans;
KW Candida stellatoides; sputum; bronchial washings; blood; milk;
KW lymph fluid; skin; soft tissue; ss.
XX
OS Synthetic.
XX
PN WO9618745-A1.
XX
PD 20-JUN-1996.
XX
PF 08-DEC-1995; 95WO-US16153.
XX
PR 16-DEC-1994; 94US-0357962.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Hoyer LL, Livi GP, Shatzman A;
XX
DR WPI; 1996-300661/30.
XX
PT C. albicans and C. stellatoides specific probes and primers - for
PT specific detection of Candida infection

```
XX
PS   Claim 1; Figure 5B; 33pp; English.
XX
CC   Five synthetic sequences (AAT29063-67) are used as probes to detect
CC   the presence of Candida albicans and Candida stellatoides in
CC   biological fluids e.g. sputum, bronchial washings, blood, milk and
CC   lymph fluid or in tissue samples e.g. skin and soft tissues.  The
CC   method of detection may also comprise using two of the probe
CC   sequences (AAT29065, AAT29066) as amplification primers and then
CC   contacting one of the probe sequences with the amplified product and
CC   detecting hybridisation.
XX
SQ   Sequence 424 BP; 126 A; 128 C; 67 G; 103 T; 0 other;

  Query Match             9.5%;  Score 359.2;  DB 17;  Length 424;
  Best Local Similarity   93.1%;  Pred. No. 4.6e-73;
  Matches  376;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       1980 GTCTCAATCATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGT 2039
             ||| |||||||||||| |||||||| ||  | || || ||||| ||||  ||||| || ||
Db         21 GTCACAATCATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGT 80

Qy       2040 TCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATA 2099
             | | || ||||||||||||||||||||||||| |||||||||||||||||||| || ||
Db         81 TATCATTAGAGAGCCACCAAACCATACTGTCACCACTACTGAATACTGGTCTCAGTCCTA 140

Qy       2100 TGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACCGATACCGTTCTTATCAGAGA 2159
             ||| || || || ||  |||| || ||||||||||||||||||||||||||||||||||||
Db        141 TGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACCGATACCGTTCTTATCAGAGA 200

Qy       2160 GCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCAC 2219
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 GCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCAC 260

Qy       2220 CACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGAGCCACCAAATCC 2279
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        261 CACTGTTACTGCACCACCAGGAGGTACCGATACTGTTATCATTAGAGAGCCACCAAATCC 320

Qy       2280 AACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACAGTTACTGC 2339
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 AACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACAGTTACTGC 380

Qy       2340 TCCTCCAGGTGGTACTGACACTGTGATTATCTATGAAAGCATGT 2383
             |||||||||||||||||||||||||||||||||||||||||||
Db        381 TCCTCCAGGTGGTACTGACACTGTGATTATCTATGAAAGCATGT 424


RESULT 2
AAT32337
ID   AAT32337 standard; DNA; 108 BP.
XX
AC   AAT32337;
XX
DT   29-NOV-1996  (first entry)
XX
DE   Probe for Candida albicans and Candida stellatoides.
XX
KW   Probe; primer; detection; identification; Candida albicans;
KW   Candida stellatoides; sputum; bronchial washings; blood; milk;
KW   lymph fluid; skin; soft tissue; ss.
XX
OS   Synthetic.
XX
PN   WO9618745-A1.
XX
PD   20-JUN-1996.
XX
PF   08-DEC-1995;   95WO-US16153.
XX
PR   16-DEC-1994;   94US-0357962.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Hoyer LL,  Livi GP,  Shatzman A;
XX
DR   WPI; 1996-300661/30.
XX
PT   C. albicans and C. stellatoides specific probes and primers - for
PT   specific detection of Candida infection
XX
PS   Claim 1; Figure 5A; 33pp; English.
XX
CC   Five synthetic sequences (AAT29063-67) are used as probes to detect
CC   the presence of Candida albicans and Candida stellatoides in
CC   biological fluids e.g. sputum, bronchial washings, blood, milk and
CC   lymph fluid or in tissue samples e.g. skin and soft tissues.  The
CC   method of detection may also comprise using two of the probe
CC   sequences (AAT29065, AAT29066) as amplification primers and then
CC   contacting one of the probe sequences with the amplified product and
CC   detecting hybridisation.
XX
SQ   Sequence 108 BP; 33 A; 32 C; 16 G; 27 T; 0 other;

  Query Match             2.9%;  Score 108;  DB 17;  Length 108;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-15;
  Matches  108;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2164 CCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACCACT 2223
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACCACT 60

Qy       2224 GTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGAGCCA 2271
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGAGCCA 108


RESULT 3
AAT32330
ID   AAT32330 standard; DNA; 108 BP.
XX
AC   AAT32330;
XX
DT   29-NOV-1996  (first entry)
XX
DE   Probe for Candida albicans and Candida stellatoides.
XX
KW   Probe; primer; detection; identification; Candida albicans;
KW   Candida stellatoides; sputum; bronchial washings; blood; milk;
KW   lymph fluid; skin; soft tissue; ss.
XX
OS   Synthetic.
XX
PN   WO9618745-A1.
XX
PD   20-JUN-1996.
XX
PF   08-DEC-1995;   95WO-US16153.
XX
PR   16-DEC-1994;   94US-0357962.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Hoyer LL,  Livi GP,  Shatzman A;
XX
DR   WPI; 1996-300661/30.
XX
PT   C. albicans and C. stellatoides specific probes and primers - for
PT   specific detection of Candida infection
XX
PS   Claim 1; Figure 5A; 33pp; English.
XX
CC   Five synthetic sequences (AAT29063-67) are used as probes to detect
CC   the presence of Candida albicans and Candida stellatoides in
CC   biological fluids e.g. sputum, bronchial washings, blood, milk and
CC   lymph fluid or in tissue samples e.g. skin and soft tissues.  The
CC   method of detection may also comprise using two of the probe
```

```
CC   sequences (AAT29065, AAT29066) as amplification primers and then
CC   contacting one of the probe sequences with the amplified product and
CC   detecting hybridisation.
XX
SQ   Sequence 108 BP; 31 A; 30 C; 15 G; 32 T; 0 other;

  Query Match             2.9%;  Score 108;  DB 17;  Length 108;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-15;
  Matches  108;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1408 CCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACT 1467
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACT 60

Qy       1468 GTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCA 1515
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCA 108


RESULT 4
AAT32331
ID   AAT32331 standard; DNA; 108 BP.
XX
AC   AAT32331;
XX
DT   29-NOV-1996  (first entry)
XX
DE   Probe for Candida albicans and Candida stellatoides.
XX
KW   Probe; primer; detection; identification; Candida albicans;
KW   Candida stellatoides; sputum; bronchial washings; blood; milk;
KW   lymph fluid; skin; soft tissue; ss.
XX
OS   Synthetic.
XX
PN   WO9618745-A1.
XX
PD   20-JUN-1996.
XX
PF   08-DEC-1995;   95WO-US16153.
XX
PR   16-DEC-1994;   94US-0357962.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Hoyer LL,  Livi GP,  Shatzman A;
XX
DR   WPI; 1996-300661/30.
XX
PT   C. albicans and C. stellatoides specific probes and primers - for
PT   specific detection of Candida infection
XX
PS   Claim 1; Figure 5A; 33pp; English.
XX
CC   Five synthetic sequences (AAT29063-67) are used as probes to detect
CC   the presence of Candida albicans and Candida stellatoides in
CC   biological fluids e.g. sputum, bronchial washings, blood, milk and
CC   lymph fluid or in tissue samples e.g. skin and soft tissues.  The
CC   method of detection may also comprise using two of the probe
CC   sequences (AAT29065, AAT29066) as amplification primers and then
CC   contacting one of the probe sequences with the amplified product and
CC   detecting hybridisation.
XX
SQ   Sequence 108 BP; 30 A; 31 C; 16 G; 31 T; 0 other;

  Query Match             2.9%;  Score 108;  DB 17;  Length 108;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-15;
  Matches  108;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1516 CCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACA 1575
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACA 60

Qy       1576 GTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCT 1623
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCT 108


RESULT 5
AAT32332
ID   AAT32332 standard; DNA; 108 BP.
XX
AC   AAT32332;
XX
DT   29-NOV-1996  (first entry)
XX
DE   Probe for Candida albicans and Candida stellatoides.
XX
KW   Probe; primer; detection; identification; Candida albicans;
KW   Candida stellatoides; sputum; bronchial washings; blood; milk;
KW   lymph fluid; skin; soft tissue; ss.
XX
OS   Synthetic.
XX
PN   WO9618745-A1.
XX
PD   20-JUN-1996.
XX
PF   08-DEC-1995;   95WO-US16153.
XX
PR   16-DEC-1994;   94US-0357962.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Hoyer LL,  Livi GP,  Shatzman A;
XX
DR   WPI; 1996-300661/30.
XX
PT   C. albicans and C. stellatoides specific probes and primers - for
PT   specific detection of Candida infection
XX
PS   Claim 1; Figure 5A; 33pp; English.
XX
CC   Five synthetic sequences (AAT29063-67) are used as probes to detect
CC   the presence of Candida albicans and Candida stellatoides in
CC   biological fluids e.g. sputum, bronchial washings, blood, milk and
CC   lymph fluid or in tissue samples e.g. skin and soft tissues.  The
CC   method of detection may also comprise using two of the probe
CC   sequences (AAT29065, AAT29066) as amplification primers and then
CC   contacting one of the probe sequences with the amplified product and
CC   detecting hybridisation.
XX
SQ   Sequence 108 BP; 34 A; 36 C; 16 G; 22 T; 0 other;

  Query Match             2.9%;  Score 108;  DB 17;  Length 108;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-15;
  Matches  108;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1624 CCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACTACT 1683
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACTACT 60

Qy       1684 GTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAACCA 1731
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAACCA 108


RESULT 6
AAT32334
ID   AAT32334 standard; DNA; 108 BP.
XX
AC   AAT32334;
XX
DT   29-NOV-1996  (first entry)
XX
DE   Probe for Candida albicans and Candida stellatoides.
```

```
XX
KW   Probe; primer; detection; identification; Candida albicans;
KW   Candida stellatoides; sputum; bronchial washings; blood; milk;
KW   lymph fluid; skin; soft tissue; ss.
XX
OS   Synthetic.
XX
PN   WO9618745-A1.
XX
PD   20-JUN-1996.
XX
PF   08-DEC-1995;   95WO-US16153.
XX
PR   16-DEC-1994;   94US-0357962.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Hoyer LL,  Livi GP,  Shatzman A;
XX
DR   WPI; 1996-300661/30.
XX
PT   C. albicans and C. stellatoides specific probes and primers - for
PT   specific detection of Candida infection
XX
PS   Claim 1; Figure 5A; 33pp; English.
XX
CC   Five synthetic sequences (AAT29063-67) are used as probes to detect
CC   the presence of Candida albicans and Candida stellatoides in
CC   biological fluids e.g. sputum, bronchial washings, blood, milk and
CC   lymph fluid or in tissue samples e.g. skin and soft tissues.  The
CC   method of detection may also comprise using two of the probe
CC   sequences (AAT29065, AAT29066) as amplification primers and then
CC   contacting one of the probe sequences with the amplified product and
CC   detecting hybridisation.
XX
SQ   Sequence 108 BP; 31 A; 30 C; 18 G; 29 T; 0 other;

  Query Match             2.9%;  Score 108;  DB 17;  Length 108;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-15;
  Matches  108;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1840 CCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACT 1899
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACT 60

Qy       1900 GTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCA 1947
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCA 108


RESULT 7
AAT32336
ID   AAT32336 standard; DNA; 108 BP.
XX
AC   AAT32336;
XX
DT   29-NOV-1996  (first entry)
XX
DE   Probe for Candida albicans and Candida stellatoides.
XX
KW   Probe; primer; detection; identification; Candida albicans;
KW   Candida stellatoides; sputum; bronchial washings; blood; milk;
KW   lymph fluid; skin; soft tissue; ss.
XX
OS   Synthetic.
XX
PN   WO9618745-A1.
XX
PD   20-JUN-1996.
XX
PF   08-DEC-1995;   95WO-US16153.
XX
PR   16-DEC-1994;   94US-0357962.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Hoyer LL,  Livi GP,  Shatzman A;
XX
DR   WPI; 1996-300661/30.
XX
PT   C. albicans and C. stellatoides specific probes and primers - for
PT   specific detection of Candida infection
XX
PS   Claim 1; Figure 5A; 33pp; English.
XX
CC   Five synthetic sequences (AAT29063-67) are used as probes to detect
CC   the presence of Candida albicans and Candida stellatoides in
CC   biological fluids e.g. sputum, bronchial washings, blood, milk and
CC   lymph fluid or in tissue samples e.g. skin and soft tissues.  The
CC   method of detection may also comprise using two of the probe
CC   sequences (AAT29065, AAT29066) as amplification primers and then
CC   contacting one of the probe sequences with the amplified product and
CC   detecting hybridisation.
XX
SQ   Sequence 108 BP; 33 A; 34 C; 15 G; 26 T; 0 other;

  Query Match             2.9%;  Score 108;  DB 17;  Length 108;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-15;
  Matches  108;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2056 CCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACCACT 2115
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACCACT 60

Qy       2116 GTTACTGCACCACCTGGTGAAACCGATACCGTTCTTATCAGAGAGCCA 2163
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTACTGCACCACCTGGTGAAACCGATACCGTTCTTATCAGAGAGCCA 108


RESULT 8
AAT29063
ID   AAT29063 standard; DNA; 108 BP.
XX
AC   AAT29063;
XX
DT   28-NOV-1996  (first entry)
XX
DE   Probe for Candida albicans and Candida stellatoides.
XX
KW   Probe; primer; detection; identification; Candida albicans;
KW   Candida stellatoides; sputum; bronchial washings; blood; milk;
KW   lymph fluid; skin; soft tissue; ss.
XX
OS   Synthetic.
XX
PN   WO9618745-A1.
XX
PD   20-JUN-1996.
XX
PF   08-DEC-1995;   95WO-US16153.
XX
PR   16-DEC-1994;   94US-0357962.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Hoyer LL,  Livi GP,  Shatzman A;
XX
DR   WPI; 1996-300661/30.
XX
PT   C. albicans and C. stellatoides specific probes and primers - for
PT   specific detection of Candida infection
XX
PS   Claim 1; Figure 5A; 33pp; English.
XX
CC   Five synthetic sequences (AAT29063-67) are used as probes to detect
CC   the presence of Candida albicans and Candida stellatoides in
```

```
CC   biological fluids e.g. sputum, bronchial washings, blood, milk and
CC   lymph fluid or in tissue samples e.g. skin and soft tissues.  The
CC   method of detection may also comprise using two of the probe
CC   sequences (AAT29065, AAT29066) as amplification primers and then
CC   contacting one of the probe sequences with the amplified product and
CC   detecting hybridisation.
XX
SQ   Sequence 108 BP; 29 A; 29 C; 19 G; 31 T; 0 other;

  Query Match             2.9%;  Score 108;  DB 17;  Length 108;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-15;
  Matches  108;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1300 CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACA 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACA 60

Qy       1360 GTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCA 1407
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCA 108


RESULT 9
AAT32333
ID   AAT32333 standard; DNA; 108 BP.
XX
AC   AAT32333;
XX
DT   29-NOV-1996  (first entry)
XX
DE   Probe for Candida albicans and Candida stellatoides.
XX
KW   Probe; primer; detection; identification; Candida albicans;
KW   Candida stellatoides; sputum; bronchial washings; blood; milk;
KW   lymph fluid; skin; soft tissue; ss.
XX
OS   Synthetic.
XX
PN   WO9618745-A1.
XX
PD   20-JUN-1996.
XX
PF   08-DEC-1995;   95WO-US16153.
XX
PR   16-DEC-1994;   94US-0357962.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Hoyer LL,  Livi GP,  Shatzman A;
XX
DR   WPI; 1996-300661/30.
XX
PT   C. albicans and C. stellatoides specific probes and primers - for
PT   specific detection of Candida infection
XX
PS   Claim 1; Figure 5A; 33pp; English.
XX
CC   Five synthetic sequences (AAT29063-67) are used as probes to detect
CC   the presence of Candida albicans and Candida stellatoides in
CC   biological fluids e.g. sputum, bronchial washings, blood, milk and
CC   lymph fluid or in tissue samples e.g. skin and soft tissues.  The
CC   method of detection may also comprise using two of the probe
CC   sequences (AAT29065, AAT29066) as amplification primers and then
CC   contacting one of the probe sequences with the amplified product and
CC   detecting hybridisation.
XX
SQ   Sequence 108 BP; 34 A; 33 C; 16 G; 25 T; 0 other;

  Query Match             2.8%;  Score 106.4;  DB 17;  Length 108;
  Best Local Similarity   99.1%;  Pred. No. 5.5e-15;
  Matches  107;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1732 CCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACCACTACCACT 1791
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACCACTACCACT 60

Qy       1792 GTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCA 1839
             ||||||||||||||||||||||||||||||||||| ||||||||||||
Db         61 GTAACTGCACCACCAGGTGGTACTGACACTGTTATTATTAGAGAGCCA 108


RESULT 10
AAT32335
ID   AAT32335 standard; DNA; 108 BP.
XX
AC   AAT32335;
XX
DT   29-NOV-1996  (first entry)
XX
DE   Probe for Candida albicans and Candida stellatoides.
XX
KW   Probe; primer; detection; identification; Candida albicans;
KW   Candida stellatoides; sputum; bronchial washings; blood; milk;
KW   lymph fluid; skin; soft tissue; ss.
XX
OS   Synthetic.
XX
PN   WO9618745-A1.
XX
PD   20-JUN-1996.
XX
PF   08-DEC-1995;   95WO-US16153.
XX
PR   16-DEC-1994;   94US-0357962.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Hoyer LL,  Livi GP,  Shatzman A;
XX
DR   WPI; 1996-300661/30.
XX
PT   C. albicans and C. stellatoides specific probes and primers - for
PT   specific detection of Candida infection
XX
PS   Claim 1; Figure 5A; 33pp; English.
XX
CC   Five synthetic sequences (AAT29063-67) are used as probes to detect
CC   the presence of Candida albicans and Candida stellatoides in
CC   biological fluids e.g. sputum, bronchial washings, blood, milk and
CC   lymph fluid or in tissue samples e.g. skin and soft tissues.  The
CC   method of detection may also comprise using two of the probe
CC   sequences (AAT29065, AAT29066) as amplification primers and then
CC   contacting one of the probe sequences with the amplified product and
CC   detecting hybridisation.
XX
SQ   Sequence 108 BP; 33 A; 37 C; 14 G; 24 T; 0 other;

  Query Match             2.8%;  Score 106.4;  DB 17;  Length 108;
  Best Local Similarity   99.1%;  Pred. No. 5.5e-15;
  Matches  107;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1948 CCAAACCCAACTGTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACC 2007
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAACCCAACTGTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACC 60

Qy       2008 ATTACCGCTCCACCTGGTGAAACTGATACCGTTCTTATCAGAGAGCCA 2055
             ||||||||||||||||||||||||||||||||||| ||||||||||||
Db         61 ATTACCGCTCCACCTGGTGAAACTGATACCGTTCTCATCAGAGAGCCA 108


RESULT 11
AAT32338
ID   AAT32338 standard; DNA; 99 BP.
XX
AC   AAT32338;
XX
```

```
DT   29-NOV-1996  (first entry)
XX
DE   Probe for Candida albicans and Candida stellatoides.
XX
KW   Probe; primer; detection; identification; Candida albicans;
KW   Candida stellatoides; sputum; bronchial washings; blood; milk;
KW   lymph fluid; skin; soft tissue; ss.
XX
OS   Synthetic.
XX
PN   WO9618745-A1.
XX
PD   20-JUN-1996.
XX
PF   08-DEC-1995;   95WO-US16153.
XX
PR   16-DEC-1994;   94US-0357962.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Hoyer LL,  Livi GP,  Shatzman A;
XX
DR   WPI; 1996-300661/30.
XX
PT   C. albicans and C. stellatoides specific probes and primers - for
PT   specific detection of Candida infection
XX
PS   Claim 1; Figure 5A; 33pp; English.
XX
CC   Five synthetic sequences (AAT29063-67) are used as probes to detect
CC   the presence of Candida albicans and Candida stellatoides in
CC   biological fluids e.g. sputum, bronchial washings, blood, milk and
CC   lymph fluid or in tissue samples e.g. skin and soft tissues.  The
CC   method of detection may also comprise using two of the probe
CC   sequences (AAT29065, AAT29066) as amplification primers and then
CC   contacting one of the probe sequences with the amplified product and
CC   detecting hybridisation.
XX
SQ   Sequence 99 BP; 29 A; 29 C; 14 G; 27 T; 0 other;

  Query Match             2.6%;  Score 99;  DB 17;  Length 99;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-13;
  Matches   99;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2272 CCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACA 2331
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACA 60

Qy       2332 GTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 2370
             |||||||||||||||||||||||||||||||||||||||
Db         61 GTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 99


RESULT 12
AAA88864/c
ID   AAA88864 standard; DNA; 8201 BP.
XX
AC   AAA88864;
XX
DT   19-FEB-2001  (first entry)
XX
DE   Human dentin sialophosphoprotein gene.
XX
KW   Dentin sialophosphoprotein; DSPP; human; SIBLINGS;
KW   integrin-binding ligand; diagnosis; antiinflammatory; therapy;
KW   osteoporosis; chromosome 4; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2387..7896
FT                   /*tag= a
FT                   /note= "contains exons"
FT   exon            2387..2437
FT                   /*tag= b
FT   intron          2438..3576
FT                   /*tag= c
FT   exon            3577..3660
FT                   /*tag= d
FT   intron          3661..3793
FT                   /*tag= e
FT   exon            3794..4780
FT                   /*tag= f
FT   intron          4781..5256
FT                   /*tag= g
FT   exon            5257..7896
FT                   /*tag= h
XX
PN   WO200062065-A1.
XX
PD   19-OCT-2000.
XX
PF   07-APR-2000; 2000WO-US09349.
XX
PR   09-APR-1999;   99US-0128468.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Fisher LW,  Fedarko NS,  Young MF;
XX
DR   WPI; 2000-679515/66.
DR   P-PSDB; AAB19772.
XX
PT   Detecting small integrin-binding ligand N-linked glycoproteins for
PT   detection of a tumor or protection against a complement mediated immune
PT   response, comprises detection where Factor H is not an inhibitor -
XX
PS   Disclosure; Page 98-100; 110pp; English.
XX
CC   The present sequence is that of DNA encoding human dentin
CC   sialophosphoprotein (DSPP, see AAB19772), a member of the small
CC   integrin binding ligand, N-linked glycoproteins (SIBLINGS) family.
CC   The invention provides methods and compositions for exploiting the
CC   discovery that members of the SIBLINGS family bind to complement
CC   Factor H, conferring resistance to complement mediated lysis.  A
CC   claimed method of conferring protection against a complement
CC   mediated immune response involves providing a reservoir or other
CC   supply in the subject's body so that a SIBLINGS protein can be
CC   dispersed to interfere with complement mediated lysis and
CC   inflammation.  This protects cells that are grafted onto foreign
CC   tissue or bone marrow cells introduced into a foreign host.  The
CC   SIBLINGS protein can be BSP, OPN, DMP1 or DSPP.  A method of
CC   detecting a SIBLINGS protein in a sample from a subject suspected
CC   of having abnormal bone turnover, especially osteoporosis, is also
CC   claimed.
XX
SQ   Sequence 8201 BP; 2801 A; 1477 C; 1873 G; 2050 T; 0 other;

  Query Match             2.5%;  Score 94.2;  DB 21;  Length 8201;
  Best Local Similarity   41.7%;  Pred. No. 1.6e-11;
  Matches  588;  Conservative    0;  Mismatches  823;  Indels    0;  Gaps    0;

Qy        952 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 1011
              | || || |   |||  | |  ||||| || ||  ||  |||.  | | |      |
Db       7290 ACTGCTGTCACTGCTATTGCTGCTGTCACTGCTGTCACTGCTATCGCTGCTGTCACTGCT 7231

Qy       1012 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1071
              | | ||||||||||| || |  |         | | | ||  |   | ||   |  ||
Db       7230 ATCGCTGCTGTCACTGCTGTTGCTGCTGTCACTACTATTACTGCTATCACTGCTGTCACT 7171

Qy       1072 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1131
                | |     ||||  |  |   | |    |  ||| |      || ||| ||| ||  |
Db       7170 GCTGTCACTGCTATCACTGCTGTTGCTGCTATCACTGCTGTCACTGCTGTCACTGCTATC 7111

Qy       1132 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1191
                |      |   ||    |   || ||   | |  |||| | | | || | |  |
```

```
Db       7110 ACTGCTGTCACTGCTATCGCTGCTGTTGCTGCTGTCACTGCTGTCGCTGCTATTGCTATC 7051

Qy       1192 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGT 1251
Db       7050 ACTGCTGTCACTGCTGTTGCTGCTGTCACCGCTATTGCTGCTGTCACTACTATTACTGCT 6991

Qy       1252 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACT 1311
Db       6990 GTCACTGCTATCACTGCTGTCACTGCTGTTGCTGCTGTCACTGCTATTGCTGTTGTCACT 6931

Qy       1312 GTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCT 1371
Db       6930 ACTATTACTGCTTTCACTGCTATCACTGCTGTTGCTGCTATCACTGCTGTCACTGCTGTT 6871

Qy       1372 CCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACTACTACT 1431
Db       6870 GCTGCTATCACTGCTGTCACTATCACTGCTGTTGCTGCTATCACTGCTGTCACTGCTATT 6811

Qy       1432 GAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGTACT 1491
Db       6810 GCTGCTTTCGTTGCTGTCACTGCTGTCACTACTATCACTGCTGTCACTGCTGTTGCTGCT 6751

Qy       1492 GACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTATTGGTCT 1551
Db       6750 GTCACTGCTATCGCTGCTGTCACTACTATTACTACTGTCACTTCTATTGCTGCTGTCGCT 6691

Qy       1552 CAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATT 1611
Db       6690 ATCACTGCTGTCACTGCTATCGCTGCTGTCGCTGCTATCACTGCTGTTGCTGCTATCGCT 6631

Qy       1612 ATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAATCTTACGCA 1671
Db       6630 GCTGTCACTACTATTACTGCTGTCACTGCTATTGCTGCTGTCGCTATCACTGCTGTCACT 6571

Qy       1672 ACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAACCA 1731
Db       6570 GCTATCACTGCTATCACTGCTGTTGCTGCTATCACTGCTGTTGCTGCTATCACTGCTGTC 6511

Qy       1732 CCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACCACTACCACT 1791
Db       6510 ACTGCTATTGCTGCTGTCGTTGCTATCACTGCTGTTGCTGCTGTCACTACTATCACTGCT 6451

Qy       1792 GTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCACCAAACCACACT 1851
Db       6450 GTCACTGCTATCACTGCTGTCACTGCTGTTGCTGCTGTCACTGCTATCGCTACTGTCACT 6391

Qy       1852 GTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGTCCA 1911
Db       6390 GCTGTTGCTGCTGTCACTGCTATCACTGCTGTCACTGCTATCGCTGCTGTTGCTGCTGTC 6331

Qy       1912 CCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACTACT 1971
Db       6330 ACTATCGCTGCTGTTACTATTACTGCTGTCACTACTATCACTGCTGTCACTATCACTGCT 6271

Qy       1972 GAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACT 2031
Db       6270 GTCACTACTATCACTGCTCTCACTGCTATTGCTGCTGTCACTACTGTCACTACTATCACT 6211

Qy       2032 GATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCT 2091
Db       6210 GCTGTTGCTGCTGTCACTGCTACTGCTGCTGTCACTGCTACTGCTGCTGTCACTGCTATC 6151

Qy       2092 CAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACCGATACCGTTCTT 2151
Db       6150 ACTGCTGTCACTACTATCACTGCTGTTACTGCTATTGCTGCTGTCGCTGCTATCACTGTT 6091

Qy       2152 ATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCT 2211
Db       6090 GTCACTACTGTCACTGCTGTTGCTGTCACTGCTGTCTGACTTACTGTCACTATCACTGCT 6031

Qy       2212 ACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGAGCCA 2271
Db       6030 GTCGCTCTCTGATTTGCTGCTGTCTGACTTGCTGTCACTGCTGTCACTACTATCACTACT 5971

Qy       2272 CCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACA 2331
Db       5970 GTCACTGCTATCACTGCTGTCACTACTGTCACTGCTATTGCTGCTATCACTACTATCACT 5911

Qy       2332 GTTACTGCTCCTCCAGGTGGTACTGACACTG 2362
Db       5910 GTCACTGCTATCACTGCTGTCACTGTCACTG 5880
```

```
RESULT 13
ABQ73537/c
ID   ABQ73537 standard; DNA; 8201 BP.
XX
AC   ABQ73537;
XX
DT   03-OCT-2002  (first entry)
XX
DE   Human dentin sialophosphoprotein precursor (DSPP) gene SEQ ID NO:1.
XX
KW   Human; dentin sialophosphoprotein precursor; dentin sialophosphoprotein;
KW   DSPP; dentinogenesis imperfecta type II; deafness; auditory;
KW   chromosome 4q21; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   promoter        1..6
FT                   /*tag= a
FT   exon            7..98
FT                   /*tag= b
FT                   /number= 1
FT   intron          99..2358
FT                   /*tag= c
FT                   /number= 1
FT   exon            2359..2437
FT                   /*tag= d
FT                   /number= 2
FT   CDS             2387..7896
FT                   /*tag= e
FT                   /product= "dentin sialophosphoprotein precursor"
FT                   /note= "contains introns"
FT   sig_peptide     2387..2431
FT                   /*tag= f
FT   mat_peptide     2432..7893
FT                   /*tag= g
FT                   /product= "dentin sialophosphoprotein"
FT                   /note= "contains introns"
FT   mat_peptide     2432..5520
FT                   /*tag= h
FT                   /product= "dentin sialoprotein"
FT                   /note= "contains introns"
FT   intron          2438..3576
FT                   /*tag= i
FT                   /number= 2
FT   exon            3577..3660
FT                   /*tag= j
FT                   /number= 3
FT   intron          3661..3793
FT                   /*tag= k
FT                   /number= 3
FT   exon            3794..4780
FT                   /*tag= l
FT                   /number= 4
FT   intron          4781..5256
FT                   /*tag= m
FT                   /number= 4
FT   exon            5257..8201
FT                   /*tag= n
FT                   /number= 5
FT   mat_peptide     5521..7893
FT                   /*tag= o
FT                   /product= "phosphophoryn"
```

```
FT   misc_feature    5596..5604
FT                   /*tag= p
FT                   /note= "cell binding domain"
FT   polyA_signal    7988..7993
FT                   /*tag= q
FT   polyA_signal    8171..8176
FT                   /*tag= r
XX
PN   WO200258722-A1.
XX
PD   01-AUG-2002.
XX
PF   30-AUG-2001; 2001WO-CN01292.
XX
PR   05-SEP-2000; 2000CN-0125042.
XX
PA   (SHAN-) SHANGHAI RES CENT BIOTECHNOLOGY.
XX
PI   Kong X, Xiao S, Zhao G, Yu C, Hu L;
XX
DR   WPI; 2002-557897/59.
DR   P-PSDB; ABP51785.
XX
PT   Diagnosis of dentinogenesis imperfecta type III and its accompanying
PT   deafness using dentin sialophosphoprotein gene and encoded products -
XX
PS   Example 3; Page 23-27; 38pp; Chinese.
XX
CC   The present invention describes a method (M1) for the diagnosis of
CC   dentinogenesis imperfecta type II and/or its accompanying deafness
CC   comprising determining the dentin sialophosphoprotein (DSPP) gene, its
CC   transcript and/or protein of an individual for comparison of their
CC   sequences with the normal sequences and judging the individual to have
CC   higher risk of suffering from the disease then the normal population.
CC   Also described are: (1) treating dentinogenesis imperfecta type III
CC   and/or its accompanying deafness by administering a safe and effective
CC   dose of normal DSPP and/or DSP protein to patients; (2) drug compositions
CC   containing safe doses of DSPP and/or DSP protein; and (3) a regent kit
CC   for detecting dentinogenesis imperfecta type II and/or its accompanying
CC   deafness containing primers for specific amplification of DSPP gene or
CC   its transcript, or containing probes for binding to the mutation site.
CC   The DSPP gene and protein sequences have auditory activity. The method
CC   (M1), dentin sialophosphoprotein (DSPP) gene and DSP protein are useful
CC   for diagnosing and treating imperfecta type II and/or its accompanying
CC   deafness. The DSPP gene is located to chromosome 4q21. The present
CC   sequence represents the human DSPP gene from the present invention.
XX
SQ   Sequence 8201 BP; 2801 A; 1477 C; 1873 G; 2050 T; 0 other;
```

Query Match 2.5%; Score 94.2; DB 24; Length 8201;
Best Local Similarity 41.7%; Pred. No. 1.6e-11;
Matches 588; Conservative 0; Mismatches 823; Indels 0; Gaps 0;

```
Qy        952 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 1011
Db       7290 ACTGCTGTCACTGCTATTGCTGCTGTCACTGCTGTCACTGCTATCGCTGCTGTCACTGCT 7231

Qy       1012 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1071
Db       7230 ATCGCTGCTGTCACTGCTGTTGCTGCTGTCACTACTATTACTGCTATCACTGCTGTCACT 7171

Qy       1072 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1131
Db       7170 GCTGTCACTGCTATCACTGCTGTTGCTGCTATCACTGCTGTCACTGCTGTCACTGCTATC 7111

Qy       1132 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1191
Db       7110 ACTGCTGTCACTGCTATCGCTGCTGTTGCTGCTGTCACTGCTGTCGCTGCTATTGCTATC 7051

Qy       1192 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGT 1251
Db       7050 ACTGCTGTCACTGCTGTTGCTGCTGTCACCGCTATTGCTGCTGTCACTACTATTACTGCT 6991

Qy       1252 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACT 1311
Db       6990 GTCACTGCTATCACTGCTGTCACTGCTGTTGCTGCTGTCACTGCTATTGCTGTTGTCACT 6931

Qy       1312 GTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCT 1371
Db       6930 ACTATTACTGCTTTCACTGCTATCACTGCTGTTGCTGCTATCACTGCTGTCACTGCTGTT 6871

Qy       1372 CCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACTACTACT 1431
Db       6870 GCTGCTATCACTGCTGTCACTATCACTGCTGTTGCTGCTATCACTGCTGTCACTGCTATT 6811

Qy       1432 GAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGTACT 1491
Db       6810 GCTGCTTTCGTTGCTGTCACTGCTGTCACTACTATCACTGCTGTCACTGCTGTTGCTGCT 6751

Qy       1492 GACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTATTGGTCT 1551
Db       6750 GTCACTGCTATCGCTGCTGTCACTACTATTACTACTGTCACTTCTATTGCTGCTGTCGCT 6691

Qy       1552 CAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATT 1611
Db       6690 ATCACTGCTGTCACTGCTATCGCTGCTGTCGCTGCTATCACTGCTGTTGCTGCTATCGCT 6631

Qy       1612 ATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAATCTTACGCA 1671
Db       6630 GCTGTCACTACTATTACTGCTGTCACTGCTATTGCTGCTGTCGCTATCACTGCTGTCACT 6571

Qy       1672 ACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAACCA 1731
Db       6570 GCTATCACTGCTATCACTGCTGTTGCTGCTATCACTGCTGTTGCTGCTATCACTGCTGTC 6511

Qy       1732 CCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACCACTACCACT 1791
Db       6510 ACTGCTATTGCTGCTGTCGTTGCTATCACTGCTGTTGCTGCTGTCACTACTATCACTGCT 6451

Qy       1792 GTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCACCAAACCACACT 1851
Db       6450 GTCACTGCTATCACTGCTGTCACTGCTGTTGCTGCTGTCACTGCTATCGCTACTGTCACT 6391

Qy       1852 GTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGTCCA 1911
Db       6390 GCTGTTGCTGCTGTCACTGCTATCACTGCTGTCACTGCTATCGCTGCTGTTGCTGCTGTC 6331

Qy       1912 CCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACTACT 1971
Db       6330 ACTATCGCTGCTGTTACTATTACTGCTGTCACTACTATCACTGCTGTCACTATCACTGCT 6271

Qy       1972 GAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACT 2031
Db       6270 GTCACTACTATCACTGCTCTCACTGCTATTGCTGCTGTCACTACTGTCACTACTATCACT 6211

Qy       2032 GATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCT 2091
Db       6210 GCTGTTGCTGCTGTCACTGCTACTGCTGCTGTCACTGCTACTGCTGCTGTCACTGCTATC 6151

Qy       2092 CAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACCGATACCGTTCTT 2151
Db       6150 ACTGCTGTCACTACTATCACTGCTGTTACTGCTATTGCTGCTGTCGCTGCTATCACTGTT 6091

Qy       2152 ATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCT 2211
Db       6090 GTCACTACTGTCACTGCTGTTGCTGTCACTGCTGTCTGACTTACTGTCACTATCACTGCT 6031

Qy       2212 ACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGAGCCA 2271
Db       6030 GTCGCTCTCTGATTTGCTGCTGTCTGACTTGCTGTCACTGCTGTCACTACTATCACTACT 5971

Qy       2272 CCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACA 2331
Db       5970 GTCACTGCTATCACTGCTGTCACTACTGTCACTGCTATTGCTGCTATCACTACTATCACT 5911

Qy       2332 GTTACTGCTCCTCCAGGTGGTACTGACACTG 2362
```

```
              || ||||||    | | ||  |||| |||||
Db       5910 GTCACTGCTATCACTGCTGTCACTGTCACTG 5880


RESULT 14
AAA61847
ID   AAA61847 standard; DNA; 5511 BP.
XX
AC   AAA61847;
XX
DT   28-OCT-2000  (first entry)
XX
DE   Cryptosporidium parvum Iowa isolate GP900 ORF.
XX
KW   GP900; Iowa isolate; glycoprotein; antibody; cryptosporidiosis;
KW   competitive inhibition; attachment; invasion; ligand binding; sporozoite;
KW   merozoite; diarrhoea; protozoacide; open reading frame; ORF; ds.
XX
OS   Cryptosporidium parvum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1.5511
FT                   /*tag= a
FT                   /product= "Cryptosporidium parvum Iowa isolate GP900"
FT                   /note= "No stop codon given"
XX
PN   US6071518-A.
XX
PD   06-JUN-2000.
XX
PF   12-SEP-1997;   97US-0928361.
XX
PR   13-SEP-1996;   96US-0026062.
PR   01-JUN-1993;   93US-0071880.
PR   29-MAY-1992;   92US-0891301.
PR   03-APR-1995;   95US-0415751.
PR   14-AUG-1996;   96US-0700651.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Petersen C;
XX
DR   WPI; 2000-422065/36.
DR   P-PSDB; AAB11726.
XX
PT   New GP900 protein fragments and fusion proteins of Cryptosporidium
PT   parvum, useful for detecting the presence of the parasite, and
PT   diagnosing or treating Cryptosporidium infections by competitive
PT   inhibition of the function of GP900 -
XX
PS   Claim 16; Column 37-42; 59pp; English.
XX
CC   The invention relates to the GP900 glycoprotein of the protozoan
CC   Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion
CC   proteins comprising GP900 fragments. The invention also relates to the
CC   administration of GP900 or fragments thereof to a host to elicit anti-
CC   GP900 antibody production, and to a method of cryptosporidiosis treatment
CC   or prophylaxis comprising administration of anti-GP900 antibodies to an
CC   individual. Cryptosporidium parvum GP900 and GP900 fragments are able to
CC   competitively inhibit sporozoite or merozoite attachment or invasion, and
CC   are also useful for the generation of anti-GP900 antibodies. The
CC   antibodies also inhibit sporozoite or merozoite attachment/invasion and
CC   additionally inhibit the binding of GP900 ligands to GP900. GP900
CC   proteins, fragments and antibodies may therefore be used to treat or
CC   prevent cryptosporidiosis. Infection with Cryptosporidium is a common
CC   cause of diarrhoea in humans and causes life-threatening diarrhoea in
CC   immunocompromised persons. Cryptosporidiosis can be contracted from
CC   contaminated municipal water supplies (e.g., public swimming pools). It
CC   is also a cause of disease in animals, resulting in financial losses in
CC   agriculture. GP900 fragments, fusion proteins and antibodies may also be
CC   used for the diagnosis of Cryptosporidium parvum infections, and for the
CC   detection of the parasite in the environment. The present sequence
CC   represents the open erading frame (ORF) encoding the GP900 protein of the
CC   Iowa isolate of Cryptosporidium parvum.
XX
SQ   Sequence 5511 BP; 1964 A; 1204 C; 944 G; 1399 T; 0 other;

  Query Match             2.3%;  Score 87;  DB 21;  Length 5511;
  Best Local Similarity   46.3%;  Pred. No. 6.4e-10;
  Matches  285;  Conservative    0;  Mismatches  330;  Indels    0;  Gaps    0;

Qy        986 CTACAACTAGAACAGTTACAGACAGTACCACTGCTGTCACTACTTTACCATTCAATCCAA 1045
              | ||||| | ||||   |||   | ||| ||| ||  |||||||    |    |  |||
Db        974 CAACAACGACAACAACAACAACTACTACAACTACTACCACTACTACTACGACAACAACAA 1033

Qy       1046 GTGTTGATAAAACCAAAACAATCGAAATTTTGCAACCTATTCCAACCACTACCATCACAA 1105
              |  |  ||  || | |||||    || |       ||| | ||||||| || |  ||||
Db       1034 CTACTACTACGACAACAACAACAACAACTACTACTACTACTACAACCACAACAACTACAA 1093

Qy       1106 CTTCATATGTTGGTGTGACTACTTCCTATCTGACTAAGACTGCACCAATTGGTGAAACAG 1165
              |  |      |      || ||  |   |   |  ||  |  || |||       ||||
Db       1094 CCACGACAACTACAACCACAACCACAACTACCAAGAAACCAACAACAACAACAACAACAA 1153

Qy       1166 CTACTGTTATTGTTGATGTGCCATATCATACTACCACAACTGTTACCAGTGAATGGACAG 1225
              | |||  || |  |       ||      || || || |||  ||| | |      ||
Db       1154 CAACTACTACTACTACAACAACAACAACAACAACTACTACTACTACTACTACTACTACTA 1213

Qy       1226 GAACAATCACTACCACCACAACTCGTACCAATCCAACTGATTCAATTGACACAGTGGTGG 1285
                || |  ||||| || || ||    ||||   | ||     ||| |  ||
Db       1214 CTACTACTACTACTACTACCACAACAACCACAACCACAACCACAACTACCAAGAAACCAA 1273

Qy       1286 TACAAGTTCCACTGCCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTG 1345
              |  |    ||    |||   ||||   || ||||||          |     ||
Db       1274 CAACAACAACAACAACAACAACAACAACTACTACTACAACCACGACAACAACAACCACGA 1333

Qy       1346 CTACAACCACTACAGTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGC 1405
              | || || || |||  |||  || |          ||   ||||   | ||  |
Db       1334 CAACCACAACCACAACTACAACTACCAAGAAACCAACAACTACTACTACTACTACCACAA 1393

Qy       1406 CACCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTA 1465
              || |||    ||||  ||||| ||   |       |      |   | |||| || || |
Db       1394 CAACAACAACTACTACTACTACCACAACAACAACAACTACTACTACTACTACAACCACAA 1453

Qy       1466 CTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAA 1525
              |    ||  |  |  ||      |||  |      |   |  | |  | || ||| | | |
Db       1454 CCACAACCACAACCGCAACCACAACTACCAAGAAACCAACAACAACAACAACAACTACTA 1513

Qy       1526 CTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTC 1585
              ||   |||||||||| || |      |   | |     || | || || ||| |||| ||
Db       1514 CTACTACTACAACCAAGAAACCAACAACAACTACCACTGCCACAACAACAACTACTACTT 1573

Qy       1586 CTCCAGGTGGTACTG 1600
              ||  |  ||  | ||
Db       1574 CTGAAACTGAGAGTG 1588


RESULT 15
ABT04776
ID   ABT04776 standard; DNA; 5511 BP.
XX
AC   ABT04776;
XX
DT   27-SEP-2002  (first entry)
XX
DE   C parvum GP900 gene fragment SEQ ID NO: 2.
XX
KW   Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis;
KW   gene; ds.
XX
OS   Cryptosporidium parvum.
XX
PN   WO200194631-A1.
XX
PD   13-DEC-2001.
XX
PF   14-MAY-2001; 2001WO-US15624.
```

XX
PR 06-JUN-2000; 2000US-0588995.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Petersen C, Barnes DA, Nelson RG, Gut J;
XX
DR WPI; 2002-566447/60.
XX
PT Detecting Cryptosporidium in biological and environmental samples and
PT diagnosis of cryptosporidiosis involves, contacting the sample with
PT Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA
PT -
XX
PS Disclosure; Page 99-101; 157pp; English.
XX
CC The present invention relates to a method of detecting Cryptosporidium in
CC biological and environmental samples, and of diagnosing
CC cryptosporidiosis. This involves obtaining a sample and contacting it
CC with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or
CC RNA, or its variant, mutant or fragment. The method is also useful for
CC detecting and identifying individual Cryptosporidium isolates based on
CC the genetic characteristics, and for diagnosis of prior or concurrent
CC Cryptosporidium infection. The present sequence is a C. parvum coding
CC sequence used in the exemplification of the invention.
XX
SQ Sequence 5511 BP; 1964 A; 1204 C; 944 G; 1399 T; 0 other;

Query Match 2.3%; Score 87; DB 24; Length 5511;
Best Local Similarity 46.3%; Pred. No. 6.4e-10;
Matches 285; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

```
Qy       986 CTACAACTAGAACAGTTACAGACAGTACCACTGCTGTCACTACTTTACCATTCAATCCAA 1045
Db       974 CAACAACGACAACAACAACAACTACTACAACTACTACCACTACTACTACGACAACAACAA 1033

Qy      1046 GTGTTGATAAAACCAAAACAATCGAAATTTTGCAACCTATTCCAACCACTACCATCACAA 1105
Db      1034 CTACTACTACGACAACAACAACAACAACTACTACTACTACTACAACCACAACAACTACAA 1093

Qy      1106 CTTCATATGTTGGTGTGACTACTTCCTATCTGACTAAGACTGCACCAATTGGTGAAACAG 1165
Db      1094 CCACGACAACTACAACCACAACCACAACTACCAAGAAACCAACAACAACAACAACAACAA 1153

Qy      1166 CTACTGTTATTGTTGATGTGCCATATCATACTACCACAACTGTTACCAGTGAATGGACAG 1225
Db      1154 CAACTACTACTACTACAACAACAACAACAACAACTACTACTACTACTACTACTACTACTA 1213

Qy      1226 GAACAATCACTACCACCACAACTCGTACCAATCCAACTGATTCAATTGACACAGTGGTGG 1285
Db      1214 CTACTACTACTACTACTACCACAACAACCACAACCACAACCACAACTACCAAGAAACCAA 1273

Qy      1286 TACAAGTTCCACTGCCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTG 1345
Db      1274 CAACAACAACAACAACAACAACAACAACTACTACTACAACCACGACAACAACAACCACGA 1333

Qy      1346 CTACAACCACTACAGTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGC 1405
Db      1334 CAACCACAACCACAACTACAACTACCAAGAAACCAACAACTACTACTACTACTACCACAA 1393

Qy      1406 CACCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTA 1465
Db      1394 CAACAACAACTACTACTACTACCACAACAACAACAACTACTACTACTACTACAACCACAA 1453

Qy      1466 CTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAA 1525
Db      1454 CCACAACCACAACCGCAACCACAACTACCAAGAAACCAACAACAACAACAACAACTACTA 1513

Qy      1526 CTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTC 1585
Db      1514 CTACTACTACAACCAAGAAACCAACAACAACTACCACTGCCACAACAACAACTACTACTT 1573

Qy      1586 CTCCAGGTGGTACTG 1600
Db      1574 CTGAAACTGAGAGTG 1588
```

Search completed: June 20, 2003, 03:30:26
Job time : 532 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 03:16:39 ; Search time 3338 Seconds
(without alignments)
18369.117 Million cell updates/sec

Title: US-09-715-876-7
Perfect score: 3786
Sequence: 1 atgcttcaacaatttacatt.....tgtccttgttcatttagtga 3786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	400.6	10.6	959	17	CNS07DAZ	AL440241 T3 end of
c	2	353	9.3	893	17	CNS07D66	AL440068 T7 end of
c	3	272.8	7.2	1050	17	CNS07CMQ	AL439368 T7 end of
c	4	164.6	4.3	1047	17	CNS07D6L	AL440083 T7 end of
	5	144.4	3.8	1011	17	CNS07DAY	AL440240 T7 end of
c	6	102	2.7	735	17	CNS04NSM	AL299119 Tetraodon
	7	100.4	2.7	773	17	CNS01VTG	AL169549 Tetraodon
	8	98.4	2.6	908	17	CNS07DJI	AL440548 T7 end of
	9	97	2.6	762	17	CNS01XFI	AL171639 Tetraodon
	10	94.8	2.5	824	17	AZ185454	AZ185454 SP_1005_A
	11	91.4	2.4	501	17	FR0048173	AL444958 Fugu rubr
	12	91.2	2.4	989	17	CNS02HA4	AL197365 Tetraodon
	13	89.6	2.4	494	17	FR0048073	AL444858 Fugu rubr
	14	82.6	2.2	450	17	FR0025683	AL018519 F.rubripe
	15	81.4	2.2	641	13	BM181884	BM181884 fv51b11.y
	16	76.2	2.0	694	17	BH355163	BH355163 CH230-81P
c	17	75.6	2.0	530	17	AZ166409	AZ166409 SP_0088_A
	18	75.6	2.0	619	17	FR0047601	AL444386 Fugu rubr
	19	75	2.0	500	17	B67199	B67199 CpG0015B Cp
c	20	74	2.0	935	17	CNS033D4	AL225985 Tetraodon
	21	73.4	1.9	619	17	FR0006944	Z90754 F.rubripes
	22	72.6	1.9	644	13	BM605179	BM605179 170006870
c	23	72	1.9	485	17	AZ434703	AZ434703 1M0221C24
	24	70.4	1.9	621	17	AZ164800	AZ164800 SP_0076_B
	25	69.6	1.8	350	17	BH879665	BH879665 ht47h12.g
	26	69.6	1.8	740	10	AV973760	AV973760 AV973760
	27	69.2	1.8	788	17	AZ183942	AZ183942 SP_1002_A
	28	68.8	1.8	331	17	BH881398	BH881398 hv25c11.b
	29	68.4	1.8	933	17	AZ204694	AZ204694 SP_0100_A
c	30	68	1.8	470	17	FR0018463	AL011359 F.rubripe
c	31	67	1.8	810	17	AZ199472	AZ199472 SP_1039_B
	32	66.4	1.8	358	17	BH777606	BH777606 fzmb013f0
	33	65.6	1.7	605	17	AZ640388	AZ640388 1M0502E21
c	34	65	1.7	642	17	CI16B8	AJ226509 Ciona int
c	35	64.2	1.7	778	13	BM170064	BM170064 EST572587
	36	64	1.7	914	17	AZ205202	AZ205202 SP_0100_A
c	37	63.4	1.7	695	13	BJ388152	BJ388152 BJ388152
	38	62.4	1.6	273	17	FR0047466	AL444251 Fugu rubr
	39	62.4	1.6	619	17	FR0013713	AL004959 F.rubripe
	40	62	1.6	415	10	BE556771	BE556771 fk95c07.y
c	41	61.6	1.6	979	17	CNS06PDW	AL409242 T7 end of
	42	61.4	1.6	550	17	FR0043207	AL130699 Fugu rubr
	43	61.4	1.6	718	17	AZ972907	AZ972907 2M0246M24
	44	61.2	1.6	619	17	FR0006038	Z89848 F.rubripes
c	45	61.2	1.6	832	13	BM170146	BM170146 EST572669

ALIGNMENTS

RESULT 1
CNS07DAZ/c
LOCUS CNS07DAZ 959 bp DNA linear GSS 08-JUL-2001
DEFINITION T3 end of clone BD0AA010H12 of library BD0AA from strain CBS 94 of Candida tropicalis, genomic survey sequence.
ACCESSION AL440241
VERSION AL440241.1 GI:12223652
KEYWORDS GSS.
SOURCE Candida tropicalis.
ORGANISM Candida tropicalis
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 959)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 959)
AUTHORS Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and Dujon,B.
TITLE Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis

```
  JOURNAL   FEBS Lett. 487 (1), 91-94 (2000)
  MEDLINE   20584726
   PUBMED   11152891
REFERENCE   3  (bases 1 to 959)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web :
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .959
                     /organism="Candida tropicalis"
                     /strain="CBS 94"
                     /db_xref="taxon:5482"
                     /clone="BD0AA010H12"
                     /clone_lib="BD0AA"
                     /note="end : T3"
     misc_feature    complement(<23. .>937)
                     /note="similar to O74623 [ Agglutinin-like protein 3
                     precursor, ALS3 ]  [ Candida albicans]
                     3 putative frameshift(s)"
                     /evidence=not_experimental
BASE COUNT      245 a    145 c    292 g    253 t     24 others
ORIGIN

  Query Match           10.6%;  Score 400.6;  DB 17;  Length 959;
  Best Local Similarity   65.7%;  Pred. No. 1.5e-92;
  Matches  617;  Conservative   22;  Mismatches  276;  Indels   24;  Gaps    4;

Qy       1338 GTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTAT 1397
              |||: |||  || ||||  |||  |||    || ||||||||||| ||| ||||  ||||
Db        938 GTCYGTTGAAACTACCATCACATATACCAACCCACCAGGTGGTACTGATTCTGTTGTTAT 879

Qy       1398 CAGAGAGCCACCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTAC 1457
              :   || ||||:|||||  ||||||||:|| |||||  ||| |        ||  |||  |:
Db        878 YGTTGAACCACMAAACCCAACTGTCACYACCACTGAGTATGG------TTCTGTTGAAAY 825

Qy       1458 TACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACC 1517
              ||| |  ||   |||    || ||||||||||||| || ||  ||||:   ||||: |:
Db        824 TACCATCACATATACCAACCCACCAGGTGGTACTGATTCTGTTKTTATYGTTGAACYTCM 765

Qy       1518 AAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGT 1577
              ||||||||||||||:|| |||||||| | |||   ||   : | ||:|| |  ||
Db        764 AAATCCAACTGTCACYACCACCGAGTATGGATCTGTTTCTGAWCCAACYACCATCACCTA 705

Qy       1578 TACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGT 1637
              ||    || |||||||| |||||  || ||  |||||   ||||| ||||||||||||||
Db        704 CACCAACCCACCAGGTGGCACTGATTCTGTTGTTATCGTTGAACCACCAAACCCAACTGT 645

Qy       1638 CACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACTACTGTGACTGCTCCTCC 1697
              |||||||| || ||| |      ||||   | ||| || |  ||    ||    || ||
Db        644 CACCACCACCGAGTATGG------ATCTGTTGAAACTACCATCACATATACCAACCCACC 591

Qy       1698 AGGAGGCACTGACTCAGTAATTATCAGAGAACCACCAAACCACACTGTCACTACTACTGA 1757
              ||| |||||||| || ||  ||||:   | ||| |||||||  ||||||||||||||| ||
Db        590 AGGTGGCACTGATTCTGTTGTTATYGTTGTACCTCCAAACCCAACTGTCACTACTACCGA 531

Qy       1758 ATACTGGTCACAATCATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGA 1817
              ||  | ||    ||  :| | ||:|| | |||    ||    ||||||||||| |||||
Db        530 GTATGGATCTGTTTCTGWTCCAACYACCATCACCTACACCAACCCACCAGGTGGCACTGA 471

Qy       1818 CACTGTTATCATTAGAGAGCCACCAAACCACACTGTCACTACTACTGAGTATTGGTCTCA 1877
               ||||| | ||    || ||||||||||  |||||||| || || |||||| | ||:

Db        470 TTCTGTTGTTATCGTTGAACCACCAAACCCAACTGTCACCACCACCGAGTATGGTTCYG- 412

Qy       1878 ATCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGCACTGATACTGTTATCAT 1937
              ||| :|||||||  ||  : ||    |||||| |||||||||||| ||||| | ||
Db        411 -----TTGAWACTACCATCACCTWCACCAACCCACCAGGTGGCACTGATTCTGTTGTTAT 357

Qy       1938 TAGGGAACCACCAAACCCAACTGTCACCACTACTGAATACTGGTCTCAATCATATGCAAC 1997
              |||||:||||| ||:||||||||||| || |||||      |   ||   ||  ||
Db        356 CGTTGAACCWCCAAATCCWACTGTCACCACCACCGAATA------TGGTTCCGTTGATAC 303

Qy       1998 CACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGTTCTTATCAGAGAGCCACC 2057
              || |  ||  :||||   ||||| |||| ||||||  | ||| |||||   || ||:||
Db        302 TACCATCACATWTACCAACCCACCAGGTGGCACTGATTCTGTTGTTATCGTTGAACCWCC 243

Qy       2058 AAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACCACTGT 2117
              |||||  |||||||||||||| || ||  | |||   ||  |: | || |||| |||
Db        242 AAACCCAACTGTCACTACTACCGAGTATGGATCTGTTTCTGAWCCAACCACCATCACCTA 183

Qy       2118 TACTGCACCACCTGGTGAAACCGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGT 2177
              ||    ||||| |||| || ||| | ||| |||||   || ||||||||||. |||||
Db        182 CACCAACCCACCAGGTGGCACTGATTCTGTTGTTATCGTTGAACCACCAAACCCAACTGT 123

Qy       2178 CACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACC 2237
              ||| || ||||| ||  | |||   ||  |  | || |||| |||    ||    |||||
Db        122 CACCACCACTGAGTATGGATCTGTTTCTGAACCAACCACCATCACCTACACCAACCCACC 63

Qy       2238 AGGTGGTACCGATACTGTTATCATTAGAGAGCCACCAAA 2276
              |||||| || ||| ||||| | ||    |  || |||||
Db         62 AGGTGGCACTGATTCTGTTGTTATCGTTGTACCTCCAAA 24


RESULT 2
CNS07D66/c
LOCUS       CNS07D66                 893 bp    DNA     linear   GSS 08-JUL-2001
DEFINITION  T7 end of clone BD0AA009G07 of library BD0AA from strain CBS 94 of
            Candida tropicalis, genomic survey sequence.
ACCESSION   AL440068
VERSION     AL440068.1  GI:12223479
KEYWORDS    GSS.
SOURCE      Candida tropicalis.
  ORGANISM  Candida tropicalis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 893)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 893)
  AUTHORS   Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
            Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 16. Candida
            tropicalis
  JOURNAL   FEBS Lett. 487 (1), 91-94 (2000)
  MEDLINE   20584726
   PUBMED   11152891
REFERENCE   3  (bases 1 to 893)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web :
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
```

lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

```
FEATURES             Location/Qualifiers
     source          1. .893
                     /organism="Candida tropicalis"
                     /strain="CBS 94"
                     /db_xref="taxon:5482"
                     /clone="BD0AA009G07"
                     /clone_lib="BD0AA"
                     /note="end : T7"
     misc_feature    complement(<308. .>808)
                     /note="similar to P46590 [ Agglutinin-like protein 1
                     precursor, ALS1 ]  [ Candida albicans]"
                     /evidence=not_experimental
BASE COUNT      206 a    138 c    290 g    253 t      6 others
ORIGIN
```

Query Match 9.3%; Score 353; DB 17; Length 893;
Best Local Similarity 66.4%; Pred. No. 3.3e-80;
Matches 606; Conservative 2; Mismatches 274; Indels 31; Gaps 6;

```
Qy       1465 ACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCA 1524
              ||   |||     | ||||||||||||||  | ||  || |||  || ||||||||||||
Db        893 ACATATACCAACWCACCAGGTGGTACTGATACCGTTGTTGTCATCGACCCACCAAATCCA 834

Qy       1525 ACTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCT 1584
              || |||||||:   |      ||||     ||| |||  ||||| | ||||  |||
Db        833 ACCGTCACTASCTACTGAGTATGGT-----TCCGTTGACACTACCATTACATATACCAAC 779

Qy       1585 CCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACC 1644
              || |||||||| |||||  | ||  || |||  || || ||||||||||||||||||||
Db        778 CCACCAGGTGGCACTGATACCGTTGTTGTCATCGACCCACCAAACCCAACTGTCACCACC 719

Qy       1645 ACTGAATATTGGTCCCAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGC 1704
              || || ||| | ||    ||| |  |||||||  |  ||    ||    || |||||  ||
Db        718 ACCGAGTATGGATCTGTTTCTGAACCAACCACCATCACATATACCAACCCACCAGGTGGT 659

Qy       1705 ACTGACTCAGTAATTATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGG 1764
              |||||  | ||  || |||  || ||||||||  |  || || ||||||||||| ||  |
Db        658 ACTGATACCGTTGTTGTCATCGACCCACCAAATCCAACCGTNACTACTACTGAGTATGG- 600

Qy       1765 TCACAATCATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTT 1824
                 ||   ||  || || | |||     ||    ||||||||||| ||||| || |||
Db        599 -----TTCCGTTGAAACTACCATCACATATACCAACCCACCAGGTGGCACTGATACCGTT 545

Qy       1825 ATCATTAGAGAGCCACCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTT 1884
               |  | |  || |||||||| |  || || ||||||||||||||| |       ||  ||
Db        544 GTTGTCATCGACCCACCAAATCCAACCGTNACTACTACTGAGTATGG------TTCCGTT 491

Qy       1885 GCTACTACCACAACTGTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAA 1944
              |  |||||||  ||    ||    |||||| |||| |||||||| ||| |  | |  ||
Db        490 GAAACTACCATTACATATACCAACCCACCAGGTGGTACTGATACCGTTGTTGTCATCGAC 431

Qy       1945 CCACCAAACCCAACTGTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACT 2004
              |||||||| ||||| || ||:||||||| ||  | ||          || ||| || | |
Db        430 CCACCAAATCCAACCGTAACYACTACTGAGTATGGTTCCGT------TGAAACTACCATT 377

Qy       2005 ACCATTACCGCTCCACCTGGTGAAAC--TGATACCGTTCTTATCAGAGAGCCACCAAACC 2062
              ||   ||||   ||||| |||| |   |||||||||| || |||  || ||||||||||
Db        376 ACATATACCAACCCACCAGGTGGCAGCGTGATACCGTTGTTGTCATCGACCCACCAAACC 317

Qy       2063 ATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTG 2122
               ||||||||  || || || ||  | |||   ||  |  | || |||| |||   |||
Db        316 CAACTGTCACCACCACCGAGTATGGATCTGTTTCTGAACCAACCACCATCACATATACCA 257

Qy       2123 CACCACCTGGTGAAACCGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTA 2182
               |||||  ||||  || |||||||||  || |||  ||  ||||||||||  ||||||||  |
Db        256 ACCCACCAGGTGGCACTGATACCGTTGTTGTCATCGACCCACCAAACCCAACTGTCACCA 197

Qy       2183 CTACTGAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCAGGTG 2242
              | || |||||  | |||   ||  |  | || |||| |||   |||    ||||||||||
Db        196 CCACCGAATATGGATCTGTTTCTGAACCAACCACCATCACATATACCAACCCACCAGGTG 137

Qy       2243 GTACCGATACTGTTATCATTAGAGAGCCACCAAATCCAACAGTTACTACTACTGAATATT 2302
              |||| ||||| ||| |  | |  || ||||||||||||||| ||||||||||||| ||| .
Db        136 GTACTGATACCGTTGTTGTCATCGACCCACCAAATCCAACCGTTACTACTACTGAGTATG 77

Qy       2303 GGTCACAATCATTTGCCACAACCACCACAGTTACTGCTCCTCCAGGTGGTACTGACACTG 2362
              |      ||  |||  || |||| ||||  |||    || ||||||||||||| || |
Db         76 G------TTCCGTTGAAACTACCATCACATATACCAACCCACCAGGTGGTACTGATACCG 23

Qy       2363 TGATTATCTATGA 2375
              |  || ||   ||
Db         22 TTGTTGTCATCGA 10
```

```
RESULT 3
CNS07CMQ/c
LOCUS       CNS07CMQ                1050 bp    DNA     linear   GSS 08-JUL-2001
DEFINITION  T7 end of clone BD0AA004B02 of library BD0AA from strain CBS 94 of
            Candida tropicalis, genomic survey sequence.
ACCESSION   AL439368
VERSION     AL439368.1  GI:12222781
KEYWORDS    GSS.
SOURCE      Candida tropicalis.
  ORGANISM  Candida tropicalis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 1050)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 1050)
  AUTHORS   Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
            Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 16. Candida
            tropicalis
  JOURNAL   FEBS Lett. 487 (1), 91-94 (2000)
  MEDLINE   20584726
   PUBMED   11152891
REFERENCE   3  (bases 1 to 1050)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web :
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1050
                     /organism="Candida tropicalis"
                     /strain="CBS 94"
                     /db_xref="taxon:5482"
                     /clone="BD0AA004B02"
                     /clone_lib="BD0AA"
                     /note="end : T7"
```

```
     misc_feature    complement(<7. .>921)
                     /note="similar to O13368 [ Agglutinin-like protein ALA1
                     precursor ]  [ Candida albicans]"
                     /evidence=not_experimental
BASE COUNT      301 a    185 c    226 g    336 t      2 others
ORIGIN

  Query Match             7.2%;  Score 272.8;  DB 17;  Length 1050;
  Best Local Similarity   56.1%;  Pred. No. 2.3e-59;
  Matches  536;  Conservative    0;  Mismatches  417;  Indels    3;  Gaps    1;

Qy        979 ATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCTGTCACTACTTTACCATTC 1038
              ||||      | ||||||||||||||||||    || ||    || || ||||||||||
Db        972 ATTGAAATCATCACTAGAACAGTTACAGAAGCAACTACCTACGTTACAACTTTACCATTT 913

Qy       1039 AATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAACCTATTCCAACCACTACC 1098
              || ||  |    | | ||| | ||| || |||||| |  |||  || ||    || ||
Db        912 GATTCAGCTACCCAAAGAACAAGAACCATTGAAATTGTTAAACAAATGCCTTTAACAACA 853

Qy       1099 ATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACTAAGACTGCACCAATTGGT 1158
              || |||   |||||||| |||||||| ||      |   ||     ||     | ||||
Db        852 ATAACAGGCTCATATGTGGGTGTGACAACAAGAAGTACAACACTTCCTTTTGTACTTGGA 793

Qy       1159 GAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACCACAACTGTTACCAGTGAA 1218
              || || ||||||||  | || ||   |||||| ||| ||| ||||| || || | |
Db        792 GAGACTGCTACTGTATTGGTGGAATCGCCATACCATATTACTACAACAGTCACAACTTTT 733

Qy       1219 TGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCAACTGATTCAATTGACACA 1278
              |||   | |  |  |||| | ||    ||   ||   ||||| | || || ||||| ||
Db        732 TGGGATGAAGAATACACTTCTACTTTGACAGTTATTGATCCATCAGAATCTATTGATACT 673

Qy       1279 GTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAG 1338
              || || ||| |   ||||    | |||||||||   |    ||||     ||||  | |
Db        672 GTCGTTGTAAATCATCCA---GCTAATCCAACTACCACACTTACTAGTCTTTGGGTTGAA 616

Qy       1339 TCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATC 1398
              | | | |||||||    ||||  ||   |     |||||   |||||  ||| ||| |
Db        615 ACTTATACTACAACTTTGACAGAAACCCATGGAGTAGGTGAAGCCGATTTTGTCATTGTT 556

Qy       1399 AGAGAGCCACCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACT 1458
              |   | ||||||||  |  ||  | || ||   ||||| |||| ||  |||     ||
Db        555 ATGTATCCACCAAATCCAACGTTAACAACCTATGAATTTTGGACAGGATCAGACACTCAA 496

Qy       1459 ACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCA 1518
              | |__||| | |  | | |||||||||||  || |  |||| |  | |||  ||
Db        495 GAGCTCATTTATACGGATATTGCTGGTGGTACTGATACAATGGTTATTAACGTACCTTCA 436

Qy       1519 AATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTT 1578
              |||||||| |  ||||  || || | |||| ||       || || ||      ||   |
Db        435 AATCCAACAGAAACTATCACTGAATTTTGGACTGGCATAATTACTTCTGTGGTCACTAAT 376

Qy       1579 ACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTC 1638
              ||    |  |||       || ||  || | ||||    |  || || ||||| ||  |
Db        375 ACGAATGGTCCTCAGCAAACAGATACAATTGTTATATTTGTTCCACCTAACCCTACAATT 316

Qy       1639 ACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCA 1698
              || || |||||||  ||||   ||  ||      | ||    |||        |   |||
Db        315 ACAACTACTGAATTCTGGTTGGAAGATTTTCATTCTACTGAAACTCAAGAGCTTGGGCCA 256

Qy       1699 GGAGGCACTGACTCAGTAATTATCAGAGAACCACCAAACCACACTGTCACTACTACTGAA 1758
                |  |||||  | ||  |  |     |  || ||||| |  ||||| ||||| ||||||
Db        255 AATGAAACTGATACTGTTGTAGTTTACGTTCCTCCAAATCCAACTGTTACTACCACTGAA 196

Qy       1759 TACTGGTCACAATCATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGAC 1818
              | |||| || |        ||||  |   ||| | |||       || | |  ||  |||
Db        195 TTCTGGCCAGATAGTGTCACCACATCCGTCACGGAAACCAATGGTCCTGATAATAAAGAC 136

Qy       1819 ACTGTTATCATTAGAGAGCCACCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAA 1878
              || ||| |  ||  |  || |  || |  |||   || |  ||||| |  ||| |  |
Db        135 ACGGTTGTTGTTTATGTTCCCCACAATCCAACTACTACCAGAACTGAATTCTGGACAGAG 76

Qy       1879 TCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGCACTGATACTGTTAT 1934
                 | ||| ||||     ||||  ||   |  |||      ||||||  |||| |
```

```
Db         75 GATTATGCAACTAGGGTTACTGAGACACATGGTCCAGACCAAACTGATGTTGTTGT 20

RESULT 4
CNS07D6L/c
LOCUS       CNS07D6L                1047 bp    DNA     linear   GSS 08-JUL-2001
DEFINITION  T7 end of clone BD0AA009H06 of library BD0AA from strain CBS 94 of
            Candida tropicalis, genomic survey sequence.
ACCESSION   AL440083
VERSION     AL440083.1  GI:12223494
KEYWORDS    GSS.
SOURCE      Candida tropicalis.
  ORGANISM  Candida tropicalis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 1047)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 1047)
  AUTHORS   Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
            Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 16. Candida
            tropicalis
  JOURNAL   FEBS Lett. 487 (1), 91-94 (2000)
  MEDLINE   20584726
   PUBMED   11152891
REFERENCE   3  (bases 1 to 1047)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr -'Web :
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1047
                     /organism="Candida tropicalis"
                     /strain="CBS 94"
                     /db_xref="taxon:5482"
                     /clone="BD0AA009H06"
                     /clone_lib="BD0AA"
                     /note="end : T7"
     misc_feature    complement(<2. .>484)
                     /note="similar to O74660 [ Agglutinin-like protein 4
                     precursor, ALS4 ]  [ Candida albicans]"
                     /evidence=not_experimental
BASE COUNT      373 a    178 c    171 g    322 t      3 others
ORIGIN

  Query Match             4.3%;  Score 164.6;  DB 17;  Length 1047;
  Best Local Similarity   60.5%;  Pred. No. 2.7e-31;
  Matches  291;  Conservative    0;  Mismatches  184;  Indels    6;  Gaps    1;

Qy          2 TGCTTCAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATCA 61
              | ||   ||||   |   || ||||  |||   |   | | | |  | ||||    ||
Db        483 TTCTCATACAAGCCATTGTGCTATTAATATGGATTCAATTGGTATCTTCAAAAGAGATAT 424
```

```
Qy         62 CTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTA------TG 115
Db        423 CTGGGATTTTTATTGGTTTCGATTCATTAACATGGAATGCTGCTTCGGATTTACCCAGTG 364

Qy        116 CTTTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTA 175
Db        363 CATACCAAGGTCCACAAATTCCTACCTGGACTGCAGAACTAACATGGTTTTTGAATGGAG 304

Qy        176 CCAGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTA 235
Db        303 AATCTGCTGAACCAGGTGATACATTCACATTAATAATGCCATGTGTATTCAAGTTCATTA 244

Qy        236 CTTCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTT 295
Db        243 CAAATCAAAACACTGTTGATTTGATTGCTGATGGCACTACTTATGCGACCTGTAATCTCA 184

Qy        296 ATTCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGA 355
Db        183 ATTCTGGAGAAGAATTTACTACTTTCTCAAGTTTAAGCTGTACTGTTTCATCTACCTTGA 124

Qy        356 AATCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAA 415
Db        123 CAACATATACCCAAGCTCGTGGTACTTTACATGTTCCTTTGACATTTAATGTGGGTAGTT 64

Qy        416 CAGGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCA 475
Db         63 CGGGAACTTCTGTTTCGTTGACTGATTCTACTTGCTTCAGACCAGGTGTAAACACAGTAA 4

Qy        476 C 476
Db          3 C 3
```

RESULT 5
CNS07DAY

```
LOCUS       CNS07DAY                1011 bp    DNA     linear   GSS 08-JUL-2001
DEFINITION  T7 end of clone BD0AA010H12 of library BD0AA from strain CBS 94 of
            Candida tropicalis, genomic survey sequence.
ACCESSION   AL440240
VERSION     AL440240.1  GI:12223651
KEYWORDS    GSS.
SOURCE      Candida tropicalis.
  ORGANISM  Candida tropicalis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 1011)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 1011)
  AUTHORS   Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
            Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 16. Candida
            tropicalis
  JOURNAL   FEBS Lett. 487 (1), 91-94 (2000)
  MEDLINE   20584726
   PUBMED   11152891
REFERENCE   3  (bases 1 to 1011)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web :
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .1011
                     /organism="Candida tropicalis"
                     /strain="CBS 94"
                     /db_xref="taxon:5482"
                     /clone="BD0AA010H12"
                     /clone_lib="BD0AA"
                     /note="end : T7"
     misc_feature    <652. .>988
                     /note="similar to P46590 [ Agglutinin-like protein 1
                     precursor, ALS1 ]  [ Candida albicans]
                     1 putative frameshift(s)"
                     /evidence=not_experimental
BASE COUNT      336 a    188 c    151 g    293 t     43 others
ORIGIN
```

```
  Query Match             3.8%;  Score 144.4;  DB 17;  Length 1011;
  Best Local Similarity   62.6%;  Pred. No. 4.5e-26;
  Matches  234;  Conservative    4;  Mismatches  135;  Indels    1;  Gaps    1;

Qy          7 CAACAATTTACATTGTTATTCCTATATTTGTCAATTGCAAGTGCAAAGACAATCACTGGT 66
Db        625 CAAACAACTCTCTTATGCTGTGCATTGATTGCTACTTCAGTTGCTAAAGAAATTTCCGGT 684

Qy         67 GTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCTTTCAAAGGG 126
Db        685 GTGTTCATAGGATTTGAATCCTTAACTTGGGATAAAGCTGGTAATTATGCATATCAAGGT 744

Qy        127 CCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACCAGTGCCAAT 186
Db        745 CCTCAATATCCAACATGGAATGCCGTGCTTGATTGGTCGTTAGATGGCACAACAACATCC 804

Qy        187 CCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCACAAACA 246
Db        805 CCAGGTGATACGTTCACTTTGATCATGCCCTGTGTCTTCAAGTTCACTACGTYTGCAYTT 864

Qy        247 TCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTAT-TCTGGTGA 305
Db        865 CTGTTGGATCTGACTGCCAATGGKATGACATATGCCACATGTGATCTTCATGCCKGGTGA 924

Qy        306 AGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCATCCAT 365
Db        925 AGAGTTTACTACTTACTCTAGMTTKACGTGTACTGCTAYAGATTCGTTAAGTCTGTTCAT 984

Qy        366 TAAGGCATTTGGTA 379
Db        985 GAAGCTATGGGATA 998
```

RESULT 6
CNS04NSM/c

```
LOCUS       CNS04NSM                735 bp    DNA     linear   GSS 21-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            123M05 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL299119
VERSION     AL299119.1  GI:8038260
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
```

```
REFERENCE   1  (bases 1 to 735)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 735)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 735)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="123M05"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG123AG03LP1-end : T7"
BASE COUNT      242 a     45 c    223 g    197 t     28 others
ORIGIN

  Query Match             2.7%;  Score 102;  DB 17;  Length 735;
  Best Local Similarity   47.9%;  Pred. No. 3.8e-15;
  Matches  304;  Conservative   11;  Mismatches  318;  Indels    2;  Gaps    2;

Qy       1015 ACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAAC-AATCGAAAT 1073
              ||| |||  ||||||    ||   | | | | || |   || || ||:|| : :   | :
Db        647 ACTACTGCTACTACTAATACAGATACTACTAMTGMTACAAATACTAAWACTRCWRCTACW 588

Qy       1074 TTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTA 1133
                |           | | || || |  || | | ||     |  ||  |||||| |
Db        587 ACTAANNAANGCNGCGGCNACNACNACTACTAATACAACAACTACTGCTACTACTACTAC 528

Qy       1134 TCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCA 1193
              |   |  |  ||| |  | | |  || | | :|||||  || || |  : |  |:  |
Db        527 TGCTAAAACTACTACTGCTACTACTGCAGCTRCTACTACTACTGCTACWATTKCWGCTGC 468

Qy       1194 TACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTAC 1253
              |||||| || |||  ||| | |       || |  || |  ||| |  |  | ||||  | |
Db        467 TACTACTACTACTACTACTACTACTACTGCTACTACTACTGCTGCTGCTACTACTGC 408

Qy       1254 CAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGT 1313
                | | | |||| | | | |    | |  |         | : |     | | | | |||
Db        407 TACTACTACTGCTGCTACTACTGCTGCTGCTACTACKGCKGCTGCTACTACTACTACTAC 348

Qy       1314 TAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCC 1373
              || ||||: |      |    ||    |   |||||| || |||:|   |||| || || |
Db        347 TACTACTWTTWCTACTGCTACTACTACTACTGCTACTACTACTRCTACTACTACTACTAC 288

Qy       1374 AGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGA 1433
                |  ||| : ||||   | ||  |       |  | |     ||||   |||||| ||
Db        287 TACTACTACTRCTACTACTACTACTACTACTACTACTACTACTACTACTACTGCTAC 228

Qy       1434 ATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGTACTGA 1493
                |    |      |   | ||||||||||||||||  |||| || || |   |  | ||
Db        227 TGCTGCTACTACCACTACTACTACTACTACTACTACTACTACTACTACTACTGCTAC 168

Qy       1494 CTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCA 1553
                | |  | || ||       |  | | | | | |||   ||||| ||       |    ||
Db        167 TGCTGCTACTACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 108

Qy       1554 ATCCTTTGCTA-CTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTA 1612
```

```
                |  ||||| ||||||||||  |||||||| || |   || ||||    | |  | ||
Db        107 TACTGCTGCTATCTACTACTACTACTACTGCTGCTACTACTGCTACTACTACTGCTACTA 48

Qy       1613 TCAGAGAACCTCCAAACCCAACTGTCACCACCACT 1647
                |      || | |   | |||   || ||  ||
Db         47 CTACTACTACTACTACTACTACTACTACTGCT 13


RESULT 7
CNS01VTG
LOCUS       CNS01VTG                773 bp    DNA     linear   GSS 12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            199C24 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL169549
VERSION     AL169549.1  GI:7807606
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 773)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 773)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 773)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .773
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="199C24"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG199BB12LP1-end : T7"
BASE COUNT      210 a    265 c     69 g    184 t     45 others
ORIGIN

  Query Match             2.7%;  Score 100.4;  DB 17;  Length 773;
  Best Local Similarity   46.5%;  Pred. No. 1e-14;
  Matches  265;  Conservative   33;  Mismatches  269;  Indels    3;  Gaps    2;

Qy       1082 CTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACTA 1141
              ||| | |:||  |||| |  |   |: |:  :| |  ||  |||||| |   |   | |
Db        172 CTACTACWACTGCTACWACTAGTGCWACWACWGCTACTGCWACTACTGCTACTACTAGTG 231

Qy       1142 AGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACCA 1201
              ||| |  | | || |   || ||||||  :| || |  |    |   |  ||||||
Db        232 CTACTACTGCTACTGCTACTACTGCTACTACWAGTGCTACTACTGCTACTGCTACTACWA 291

Qy       1202 CAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCAA 1261
              |  |:  |||   :      ||    :|:|  ||||| ||  | |||  ||| | | | :
Db        292 CTGCWACTACTGCWACTGCTACTACTRCWACTACTACWACTGCTACTACTACTACTACTR 351

Qy       1262 CTGATTCAATTG--ACACAGTGGTGGTACAAGTTCCACTGCCAAATCCAAC-TGTTAGTA 1318
              ||  | | | ||  :  |||      ||| | | |.||| | | || | || |  |   |
```

```
Db        352 CTACTACTACTGCAWACCAGCATACATACTACTGCTACTACTACATGCTACATACTGCAA 411

Qy       1319 CTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTG 1378
              ||||:| |  |    ||    |   : |||| |||||:|||   ||| |: |: |::
Db        412 CTACWGCAACTGCWACTRCTACAGCWACTACTACCACWACAACAACTACWRCWACWRCAA 471

Qy       1379 GTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATATT 1438
              ||| | |||    | ||  : :    |  | |    ||||    |||||:||     |
Db        472 CTACTGCTACAACTACTACTRCWACTGCTACTACTACTACTACTGCTACTRCTACTACTA 531

Qy       1439 GGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAG 1498
                |   : |   : ||||: |||||||||| ||||:|  || |   |  :||||   |:
Db        532 CTACTACWGCTACWACTACWGCTACTACTGCTACTRCAACTGCTACTACWACTGCTACWA 591

Qy       1499 TAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCT 1558
                :|      | |  ||| | | |||:  |||:||  |  :          |||
Db        592 CTGCWACTGCTACAACTACAACTACTACTRCAACTRCATACATRCTACTACTACAACATA 651

Qy       1559 TTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAG 1618
              | || ||||| ||||   |||||| || | | |  |:||    | :  | ||  |
Db        652 CTACTGCTACTGCTACTACTACTGCTACTACTGCTACTRCTACTACTRCTACTACTACTA 711

Qy       1619 AACCTCCAAACCCAACTGTCACCACCACTG 1648
                || | |   | ||||  ||  | ||||
Db        712 CTACTGCTACTACTACTGCTACTGCTACTG 741


RESULT 8
CNS07DJI
LOCUS       CNS07DJI                908 bp    DNA     linear   GSS 08-JUL-2001
DEFINITION  T7 end of clone BD0AA013B05 of library BD0AA from strain CBS 94 of
            Candida tropicalis, genomic survey sequence.
ACCESSION   AL440548
VERSION     AL440548.1  GI:12223959
KEYWORDS    GSS.
SOURCE      Candida tropicalis.
  ORGANISM  Candida tropicalis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 908)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 908)
  AUTHORS   Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and
            Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 16. Candida
            tropicalis
  JOURNAL   FEBS Lett. 487 (1), 91-94 (2000)
  MEDLINE   20584726
   PUBMED   11152891
REFERENCE   3  (bases 1 to 908)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web :
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .908
                     /organism="Candida tropicalis"
                     /strain="CBS 94"
                     /db_xref="taxon:5482"
                     /clone="BD0AA013B05"
                     /clone_lib="BD0AA"
                     /note="end : T7"
     misc_feature    <47. .>727
                     /note="similar to O74623 [ Agglutinin-like protein 3
                     precursor, ALS3 ]  [ Candida albicans]"
                     /evidence=not_experimental
BASE COUNT      264 a    163 c    185 g    288 t      8 others
ORIGIN

  Query Match             2.6%;  Score 98.4;  DB 17;  Length 908;
  Best Local Similarity   56.0%;  Pred. No. 3.7e-14;
  Matches  186;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy       1836 GCCACCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCAC 1895
              |||||| || |  ||  |||| || || || | ||||  ||||   | |   |||||
Db         49 GCCACCGAATCCTACCATCACGACAACGGAATTTTGGGGTCAAGTATATTTCACTACAGA 108

Qy       1896 AACTGTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCC 1955
              ||||   |    |   ||| | |  ||||||||| | |||   |  ||   || ||
Db        109 AACTTACAAGAATAGTGCAAATAGTGTTGATACTGTTGTGATTTATGCTCCTGTCAATCC 168

Qy       1956 AACTGTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGC 2015
              ||| || ||    || || |  |||  | ||   | | ||||  | |  ||   |||
Db        169 AACAGTTACGTTGACAGAGTTTTGGGTTGAAGATTTTACAACTTCGATCACGTATACAAA 228

Qy       2016 TCCACCTGGTGAAACTGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTAC 2075
              || ||| |   | | |||||| ||| | |||   |  || |||||  ||||| ||
Db        229 TCAACCAGCGCATAGTGATACTGTTATAATCTACGTTCCCCCAAACCCAACTGTTACAGT 288

Qy       2076 TACTGAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGA 2135
              ||||||  | |||  |    | |||| ||    |  ||       | |  |||||| |
Db        289 CACTGAAGAATGGGTTGGGACTTATGGTAGTGTCGTTACGCAACTGGAATTACCTGGAGG 348

Qy       2136 AACCGATACCGTTCTTATCAGAGAGCCACCAA 2167
              ||| ||||| ||  ||      |  |||  ||
Db        349 AACAGATACAGTAATTGAATTTGTACCAGAAA 380


RESULT 9
CNS01XFI
LOCUS       CNS01XFI                762 bp    DNA     linear   GSS 12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            202I24 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL171639
VERSION     AL171639.1  GI:7809696
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 762)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 762)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
```

```
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 762)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .762
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="202I24"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG202BE12LP1~end : T7"
BASE COUNT       77 a    306 c    168 g    205 t      6 others
ORIGIN

  Query Match             2.6%;  Score 97;  DB 17;  Length 762;
  Best Local Similarity   46.2%;  Pred. No. 7.7e-14;
  Matches  308;  Conservative    3;  Mismatches  356;  Indels    0;  Gaps    0;

Qy       1465 ACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCA 1524
              | ||   : ||||||||||  | | ||    | |    | |    |  || |||  | |
Db         85 AATGGAGMMGCTCCTCCAGCCGCTCCTCCAGCTGCTCCTGTAGCTGCTCCTCCAGCTGCT 144

Qy       1525 ACTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCT 1584
              ||||  || |  | :  |    .  ||  | |   | || |  || || | ||  |||||
Db        145 CCTGTAGCTGCTCCTVCATCCGCTCCTGTAGCTGCTCCTCCAGCTGCTCCTGTAGCTGCT 204

Qy       1585 CCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACC 1644
              ||||||| || | |||   | |    | |    |  ||||||    |  ||||  |  |
Db        205 CCTCCAGCTGCTCCTGTAGCCGCTCCTGTAGCCGCTCCTCCAGCTGCTCCTGTAGCTGCT 264

Qy       1645 ACTGAATATTGGTCCCAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGC 1704
              ||| |  |    | | | ||    |     |  || ||    ||||||||  ||  |
Db        265 CCTGTAGCTGCTCCTCCAGCTGCTCCTGTAGCTGCTCCTCCAGCTGCTCCTGTAGCTGCT 324

Qy       1705 ACTGACTCAGTAATTATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGG 1764
              ||    | |    | |    |  || |||  |    ||||  || || ||  |
Db        325 CCTCCAGCTGCTCCTGTAGCTGCTCCTCCATCCGCTCCTGTAGCTGCTCCTCCAGCTGCT 384

Qy       1765 TCACAATCATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTT 1824
              |  | |   | |  |  || |  ||||| |||| ||    || || | ||     | | |
Db        385 CCTGTAGCTGCTCCTCCAGCTGCTCCTGTAGCTGCTCCTGTAGCTGCTCCTCCAGCCGCT 444

Qy       1825 ATCATTAGAGAGCCACCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTT 1884
              |    |  || |||       ||||  || || ||     |    ||  | | |   |
Db        445 CCTGTAGCTGCTCCTCCAGCTGCTCCTGTAGCTGCTCCTCCAGCTGCTCCTGTAGCTGCT 504

Qy       1885 GCTACTACCACAACTGTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAA 1944
              || |  || |  ||||| ||| ||| |||   |   |||   ||| |          |
Db        505 CCTCCATCCGCTCCTGTAGCTGCTCCTCCATCCGCTCCTGTAGCTGCTCCTCCATCTGCT 564

Qy       1945 CCACCAAACCCAACTGTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACT 2004
              ||    |   |  ||    || || ||| |      ||| | |   | |  |  || ||
Db        565 CCTGTAGCTGCTCCTCCATCCGCTCCTGTAGCTGCTCCTCCAGCCGCTCCTCCAGCTGCT 624

Qy       2005 ACCATTACCGCTCCACCTGGTGAAACTGATACCGTTCTTATCAGAGAGCCACCAAACCAT 2064
              |  | ||||||| || | ||   |||  ||| || |      |  ||    |      |
Db        625 CCTGTAGCCGCTCCTCCAGCTGNTCCTGTAGCCGCTCCTCCATCTGTTCCTGTAGCTGNT 684

Qy       2065 ACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCA 2124
              ||    |  || ||| |       ||| | |   | ||  | |  |  ||:|  ||||
Db        685 CCTCAATCCGCTCCTGTAGCTGCTCCTCCAGCTGCTCCTGTAGCTGCTCCTKTAGCTGCT 744

Qy       2125 CCACCTG 2131
              || || |
Db        745 CCTCCAG 751
```

```
RESULT 10
AZ185454
LOCUS       AZ185454                824 bp    DNA     linear   GSS 30-AUG-2000
DEFINITION  SP_1005_A2_H08_T7A Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=1005 Col=16 Row=O, DNA sequence.
ACCESSION   AZ185454
VERSION     AZ185454.1  GI:8357931
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 824)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 1005  row: O  column: 16
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 824.
FEATURES             Location/Qualifiers
     source          1. .824
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=1005 Col=16 Row=O"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT      230 a    223 c     99 g    271 t      1 others
ORIGIN

  Query Match             2.5%;  Score 94.8;  DB 17;  Length 824;
  Best Local Similarity   50.0%;  Pred. No. 3e-13;
  Matches  263;  Conservative    0;  Mismatches  262;  Indels    1;  Gaps    1;

Qy       1126 ACTTCCTATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTG 1185
              |||   |||||  ||| ||||    | | |  |   ||  |||||| |  ||  | ||
Db         22 ACTCTGTATCTAGCTAGGACTAACACTACTACTACTACTACTACTGCTGCTGCAGCTGCT 81

Qy       1186 CCATATCATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACA 1245
              |  ||  |||||| || |||  ||| | |       ||    || |  ||||| || ||
Db         82 GCTAATACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 141

Qy       1246 ACTCGTACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCAAAT 1305
              |||  ||| | | |  |||    |   ||  ||                  | |     | | |
Db        142 ACTACTACTACTGCTGCTGCAGCTGCTGCTACTACTACTACTACTACTACTACTACT 201

Qy       1306 CCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACT 1365
              | |||| |    ||  |||     |    ||  | |  |||| |||| ||  ||  | | | ||
Db        202 ACTACTGCTGCAGCTGCTGCTGCTACTACT-ACTACTTTACTACTACTGCTGCTGCTGCT 260

Qy       1366 GCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACT 1425
              |||  ||  | |  ||  |||   ||||   | ||      |   |  |       ||||   |||
Db        261 GCTGCTGCTGCTGCTACTACTACTACTACTACTGCTGCTGCTGCTGCTACTACTACTACT 320

Qy       1426 ACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGT 1485
              ||||||  |  |    |     |   | ||||||||||||||  |||| || ||||   |
Db        321 ACTACTACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTCCTACT 380
```

```
Qy       1486 GGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTAT 1545
              |||| |  |    | ||  |       |  | | | | |||   |||||  |      |
Db        381 AATACTAATACTACTACTACTACTTCTACTTCTACTACTACTACTACTACTTCTACTACT 440

Qy       1546 TGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCA 1605
              ||    |   | ||||||||||||||   | || || || |   |  ||||  ||
Db        441 AATACTACTACTACTACTACTACTACTACTTCTACTACTACTACTACTACTACTTCT 500

Qy       1606 GTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAAT 1651
              | || |   |||    ||       | ||| || || | | |
Db        501 ACTACTAACGAGTAACGAGTAAGTAATTTTTTCATCAACAATAATT 546


RESULT 11
FR0048173
LOCUS       FR0048173               501 bp    DNA     linear   GSS 05-JAN-2001
DEFINITION  Fugu rubripes GSS sequence, clone 264E22cA9, genomic survey
            sequence.
ACCESSION   AL444958
VERSION     AL444958.1  GI:12053458
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 501)
  AUTHORS   Clark,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of BAC (pBeloBACII) cloned
            genomic sequence
            The BACs can be obtained from http://www.incyte.com.
FEATURES             Location/Qualifiers
     source          1. .501
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone="264E22cA9"
                     /clone_lib="BAC 264E22"
BASE COUNT      132 a    161 c     35 g    173 t
ORIGIN

  Query Match             2.4%;  Score 91.4;  DB 17;  Length 501;
  Best Local Similarity   49.5%;  Pred. No. 1.7e-12;
  Matches  236;  Conservative    0;  Mismatches  241;  Indels    0;  Gaps    0;

Qy       1123 ACTACTTCCTATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGAT 1182
              |||||| |   |    |||  |||||  | | |   |   || ||||||  || || || |
Db         12 ACTACTGCTGCTGCTTCTACTACTGCTACTACTACTACTACTGCTACTACTACTGCTGCT 71

Qy       1183 GTGCCATATCATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACC 1242
                |   | |||||| || |||| ||| | |       |    || |  |||||  ||
Db         72 TCTACTACTACTACTACTACTACTGCTACTACTACTACTGCTACTACTACTACTACTACT 131

Qy       1243 ACAACTCGTACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCACTGCCA 1302
              ||  ||  | |   | | |||| | | | |   ||    |          | |    |
Db        132 ACTGCTGCTGCTTCTACTACTGCTACTACTACTACTACTGCTACTACTACTGCTGCTTCT 191

Qy       1303 AATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTT 1362
              | | | |||  || |||||||      |    ||    |   |||||| || |||||   |
Db        192 ACTACTACTACTACTACTACTACTACTACTACTACTGCTGCTACTACTACTACTACT 251

Qy       1363 ACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTC 1422
              ||| || || |   |  |||   | |||  | ||  |       |  | |    ||||
Db        252 ACTACTACTACTACTACTACTACTGCTGCTACTACTACTACTGCTACTACTACTACTACT 311

Qy       1423 ACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCA 1482
              |||||| ||      |    |      |   | ||||||||||||||  | ||||| || |
Db        312 ACTACTGCTTCTACTACTACTACTACTACTACTACTACTACTACTACTGCTGCTTCTACT 371

Qy       1483 GGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAG 1542
              || ||||    |    | ||  |       |  | | | | |||   | ||| |
Db        372 ACTGCTACTACTACTACTACTACTACTACTACTGCTACTACTACTACTAATACTGCTACT 431

Qy       1543 TATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACT 1599
              |    ||    |   | ||||||||||| |   || | || || | |   | ||| |
Db        432 ACTACTGCTGCTACTACTACTACTACTACTGCTAGTAGTACTACTACTACTAGTAGT 488


RESULT 12
CNS02HA4
LOCUS       CNS02HA4                989 bp    DNA     linear   GSS 13-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            139021 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL197365
VERSION     AL197365.1  GI:7835515
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 989)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 989)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 989)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .989
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="139021"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG139AH11LP1-end : T7"
BASE COUNT      320 a    245 c    122 g    261 t     41 others
ORIGIN

  Query Match             2.4%;  Score 91.2;  DB 17;  Length 989;
  Best Local Similarity   47.0%;  Pred. No. 2.8e-12;
  Matches  238;  Conservative    9;  Mismatches  259;  Indels    0;  Gaps    0;

Qy       3113 CTACAGGTGATAATGGAGACAATACTTCATCAACCAATCCAGTTCCAACTGTGGCAACAA 3172
              ||||   |  || |      | |||| |  | || | | |   | | |||    | || |
Db         31 CTACTACTACTACTACTACTACTACYACTACTACTACTACTACTACTAATACTACTA 90

Qy       3173 GTACTTTAGCATCTGCAAGTGAAGAAGACAACAAAAGCGGTTCTCATGAATCAGCATCCA 3232
              ||||    |  || | | |       | |  :|||    : |:  |    |  |  | |
Db         91 CTACTACTACTACTACTACTACTAATACAMAAACTACWACWACTACTACTACTACTA 150
```

```
Qy       3233 CAAGTTTGAAACCAAGTATGGGTGAAAATTCTGGATTAACTACTTCTACTGAAATTGAAG 3292
Db        151 CTAATACTACTACTAATACTACTAATACTACTACTAATACTACTACTACTAATACTACTA 210

Qy       3293 CTACAACAACCAGTCCTACAGAAGCTCCATCACCTGCTGTTTCTTCTGGTACTGATGTAA 3352
Db        211 CTACTACTACTACTACTACTACTACTACTACTRCTACTACTACTACTACTGCTG 270

Qy       3353 CTACTGAACCAACTGATACTAGAGAACAACCTACTACATTATCAACTACTTCAAAAACAA 3412
Db        271 CTAATACTACTACTAATACTAATACTACTACTACTACTACTACTACTACTACTACTACTA 330

Qy       3413 ACAGTGAACTGGTTGCTACTACACAAGCTACTAATGAAAATGGTGGTAAATCTCCATCAA 3472
Db        331 CTACTGCTACTACTACTACTACTACTACTGCTAATACTACTACTAATACTACTACTACTA 390

Qy       3473 CTGATTTAACATCAAGCTTGACAACAGGCACCTCAGCATCTACAAGTGCTAATAGCGAAC 3532
Db        391 CTACTACTACTACTACTACTACTACYACTACTACTAATAATAATAATAATAATAATAATA 450

Qy       3533 TTGTTACTAGTGGATCTGTTACTGGTGGAGCTGTTGCCAGTGCTTCAAATGATCAATCAC 3592
Db        451 ATNNTNNTNGTNGTGGTGGKGGTGGTGGTKCTDCTACTACTACTAATAATAATAATAATA 510

Qy       3593 ATTCTACTTCTGTTACCAACAGCAAC 3618
Db        511 ATAATAATAMTACTACTCCCTCCATC 536


RESULT 13
FR0048073
LOCUS       FR0048073               494 bp    DNA     linear   GSS 05-JAN-2001
DEFINITION  Fugu rubripes GSS sequence, clone 263K15bD8, genomic survey
            sequence.
ACCESSION   AL444858
VERSION     AL444858.1  GI:12052694
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 494)
  AUTHORS   Clark,M.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of BAC (pBeloBACII) cloned
            genomic sequence
            The BACs can be obtained from http://www.incyte.com.
FEATURES             Location/Qualifiers
     source          1. .494
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone="263K15bD8"
                     /clone_lib="BAC 263K15"
BASE COUNT      128 a    163 c     45 g    158 t
ORIGIN

  Query Match             2.4%;  Score 89.6;  DB 17;  Length 494;
  Best Local Similarity   49.0%;  Pred. No. 5e-12;
  Matches  239;  Conservative    0;  Mismatches  249;  Indels    0;  Gaps    0;

Qy       1102 ACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACTAAGACTGCACCAATTGGTGAA 1161
Db          6 ACGACTACTACTACTACTGCGGCTTCTACTACTACTAGTACTGCTGCTACTACTACGACT 65

Qy       1162 ACAGCTACTGTTATTGTTGATGTGCCATATCATACTACCACAACTGTTACCAGTGAATGG 1221
Db         66 ACGACTACTGCTGCGTCGACTACTGCGACGACGACTGCTACTACTACTGCGGCTGCTTCT 125

Qy       1222 ACAGGAACAATCACTACCACCACAACTCGTACCAATCCAACTGATTCAATTGACACAGTG 1281
Db        126 ACTACTGCTACTACTACGACGACTGCTACTACTACTACTGCTACTACTACTACTACTACT 185

Qy       1282 GTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCC 1341
Db        186 ACTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 245

Qy       1342 TTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGA 1401
Db        246 ACTGCTACTACTACTACTACTACTACTACTACTACTACTACTGCTACTGCTACTACTACT 305

Qy       1402 GAGCCACCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACT 1461
Db        306 GCTACTACTACTACTACTACTACTACTACTACTGCTGCTGCTACTGCTGCTACTACTACT 365

Qy       1462 ACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAAT 1521
Db        366 ACTACTACTACTGCTACTACTACTACTACTGCTACTACTACTACTACTACTACTACTACT 425

Qy       1522 CCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACT 1581
Db        426 ACTGCTGCTACTACTACTGCTACTACTACTACTACTACTACTGCTACTACTACTACTACT 485

Qy       1582 GCTCCTCC 1589
Db        486 GCTACTAC 493


RESULT 14
FR0025683
LOCUS       FR0025683               450 bp    DNA     linear   GSS 10-DEC-1997
DEFINITION  F.rubripes GSS sequence, clone 154E17aC12, genomic survey sequence.
ACCESSION   AL018519
VERSION     AL018519.1  GI:2684887
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y.,
            Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone="154E17aC12"
                     /clone_lib="cosmid 154E17"
BASE COUNT       96 a    140 c     64 g    144 t      6 others
ORIGIN

  Query Match             2.2%;  Score 82.6;  DB 17;  Length 450;
  Best Local Similarity   49.3%;  Pred. No. 3.1e-10;
  Matches  208;  Conservative    0;  Mismatches  214;  Indels    0;  Gaps    0;

Qy       1276 ACAGTGGTGGTACAAGTTCCACTGCCAAATCCAACTGTTAGTACTACTGAATATTGGTCT 1335
```

```
Db         29 ACTGGAGACNTCAAATAACCTCAGTGGCCTCTAACTCCTACTAATGCTACTACTACTACT 88

Qy       1336 CAGTCCTTTGCTACAACCACTACAGTTACTGCTCCTCCAGGTGGTACCGATACTGTGATT 1395

Db         89 ACTGCTACTGCTACTACTACTGCTACTACTACTACTACTACTACTGCTACTACT 148

Qy       1396 ATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCT 1455

Db        149 GCTACTACTACTACTACTGCTACTACTACTACTACTGCTACTACTGCTACTGCTGCTGCT 208

Qy       1456 ACTACTACTACTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCA 1515

Db        209 ACTGCTACTATTACTACTGCTACTACTGCTGCTGCTACTACTGCTACTACTACTGCTACT 268

Qy       1516 CCAAATCCAACTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACA 1575

Db        269 ACTGCTACTGCTGCTGCTACTGCTGCTGCTACTGCTGCTGCTACTGCTACTATTACTACT 328

Qy       1576 GTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCTCCAAACCCAACT 1635

Db        329 GCTACTACTGCTGCTGCTACTACTGCTACTACTACTGCTACTGCTGCTGCTGCTACTACT 388

Qy       1636 GTCACCACCACTGAATATTGGTCCCAATCTTACGCAACCACAACTACTGTGACTGCTCCT 1695

Db        389 GCTACTACTACTGCTACTGCTGCTGCTGCTACNACTGCTACNACTACNGCGACTGCTGAT 448

Qy       1696 CC 1697

Db        449 NC 450
```

RESULT 15
BM181884
LOCUS BM181884 641 bp mRNA linear EST 11-DEC-2001
DEFINITION fv51b11.y1 Sugano SJD adult male Danio rerio cDNA clone 5412044 5' similar to contains element TAR1 repetitive element ;, mRNA sequence.
ACCESSION BM181884
VERSION BM181884.1 GI:17512842
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes ; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 641)
AUTHORS Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE WashU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: T3 ET from Amersham
High quality sequence stop: 527.
FEATURES Location/Qualifiers
source 1. .641
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5412044"
/clone_lib="Sugano SJD adult male"
/sex="male"
/tissue_type="whole body"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: pME18S-FL3; Site_1: DraIII (CACCATGTG); Site_2: DraIII (CACTGTGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed and donated by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA. "
BASE COUNT 218 a 234 c 37 g 152 t
ORIGIN

Query Match 2.2%; Score 81.4; DB 13; Length 641;
Best Local Similarity 48.1%; Pred. No. 7.7e-10;
Matches 297; Conservative 0; Mismatches 311; Indels 9; Gaps 2;

```
Qy       2661 ATCCTCAATCGTAACTAGTACTGTTCCTACTGCAAGTACAATGTCTGATTCACTTTCTTC 2720

Db         34 ATCTCCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGC 93

Qy       2721 AACTGATGGTATTAGTGCTACATCTTCTGATAATGTTTCAAAATCAGGAGTATCAGTTAC 2780

Db         94 TACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCAGCTACATCAACTTCTCC 153

Qy       2781 AACCGAAACTTCTGTTACAACTATTCAAACTACTCCAAACCCATTATCATCTTCAGTGAC 2840

Db        154 AACAACAACAACTTCTCCAGCTACATCAACTTCTCCAACAACAACAACTTCTCCAACAAC 213

Qy       2841 ATCATTGACTCAGTTGTCTTCAATTCCAAGTGTTTCAGAAAGTGAAAGTAAAGTTACATT 2900

Db        214 AACAACTTCTCCAGCTACATCAACTTCAACAACAACAACAACT---TCTCCAGCTACATC 270

Qy       2901 TACAAGCAATGGAGACAACCAAAGTGGTACTCATGATTCACAATCTACTTCCACTGAAAT 2960

Db        271 AACTTCAACAACAACAACTTCTCCAGCTACAACTTCTCCAGCAACAACATCAACTTCTCC 330

Qy       2961 TGAAATTGTAACAACCAGTTCTACTAAAGTTTTACCACCTGTCGTTTCTTCTAATACTGA 3020

Db        331 AGCTACATCAACTTCTCCAACAACAACAACTTCTCCAGCTACATCAACTTCTCCAACAAC 390

Qy       3021 TTTGACTAGTGAACCAACAAATACCAGAGAACAACCAACTACATTATCAACTACTTCAAA 3080

Db        391 AACAACTTCTCCAGCTACATCAACTTCTCCAACAACAACAACTTCTCCAGCTACATCAAC 450

Qy       3081 CTCCATCACTGAAGATATCACCACATCTCAACCTACAGGTGATAATGGAGACAATACTTC 3140

Db        451 TTCAACAAC------AACAACAACTTCTCCAGCTACATCAACTTCAACAACAACAACTTC 504

Qy       3141 ATCAACCAATCCAGTTCCAACTGTGGCAACAAGTACTTTAGCATCTGCAAGTGAAGAAGA 3200

Db        505 TCCAGCTACAACTTCTCCAGCAACAACATCAACTTCTCCAGCTACATCAACTTCAACAAC 564

Qy       3201 CAACAAAAGCGGTTCTCATGAATCAGCATCCACAAGTTTGAAACCAAGTATGGGTGAAAA 3260

Db        565 AACAACAACTTCTCCAGCTACATCAACTTCAACAACAACAACAACTTCTCCAGCTACAAC 624

Qy       3261 TTCTGGATTAACTACTT 3277

Db        625 TTCTCCAGCAACAACAT 641
```

Search completed: June 20, 2003, 06:13:39
Job time : 3348 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 03:17:58 ; Search time 128 Seconds
(without alignments)
9070.924 Million cell updates/sec

Title: US-09-715-876-7
Perfect score: 3786
Sequence: 1 atgcttcaacaatttacatt.....tgtccttgttcatttagtga 3786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1069.4	28.2	1071	1	US-08-357-962-1	Sequence 1, Appli
	2	1069.4	28.2	1071	1	US-08-878-106-1	Sequence 1, Appli
	3	348.2	9.2	425	1	US-08-357-962-5	Sequence 5, Appli
	4	348.2	9.2	425	1	US-08-878-106-5	Sequence 5, Appli
	5	87	2.3	5511	3	US-08-928-361B-2	Sequence 2, Appli
	6	87	2.3	7334	3	US-08-928-361B-1	Sequence 1, Appli
	7	86.2	2.3	5163	3	US-08-700-651-1	Sequence 1, Appli
	8	86.2	2.3	5163	3	US-08-928-361B-4	Sequence 4, Appli
	9	86.2	2.3	5318	3	US-08-700-651-2	Sequence 2, Appli
	10	86.2	2.3	5318	3	US-08-928-361B-3	Sequence 3, Appli
	11	85.2	2.3	108	1	US-08-357-962-2	Sequence 2, Appli
	12	85.2	2.3	108	1	US-08-878-106-2	Sequence 2, Appli
c	13	82.4	2.2	100	1	US-08-145-705A-32	Sequence 32, Appl
c	14	82.4	2.2	100	1	US-08-145-705A-34	Sequence 34, Appl
c	15	76	2.0	100	1	US-08-145-705A-33	Sequence 33, Appl
c	16	68	1.8	100	1	US-08-145-705A-36	Sequence 36, Appl
	17	58	1.5	3168	4	US-09-165-239A-3	Sequence 3, Appli
c	18	57.2	1.5	100	1	US-08-145-705A-37	Sequence 37, Appl
c	19	57	1.5	100	1	US-08-145-705A-35	Sequence 35, Appl
c	20	54.2	1.4	3489	2	US-08-728-323A-1	Sequence 1, Appli
c	21	54.2	1.4	3489	4	US-09-298-568-1	Sequence 1, Appli
	22	54.2	1.4	32207	2	US-08-770-379-20	Sequence 20, Appl
	23	54.2	1.4	32207	4	US-08-757-669A-20	Sequence 20, Appl
	24	54.2	1.4	32207	4	US-09-230-371A-20	Sequence 20, Appl
	25	50.2	1.3	4197	2	US-08-682-517-7	Sequence 7, Appli
	26	50.2	1.3	4197	2	US-08-682-517-8	Sequence 8, Appli
c	27	49	1.3	3337	1	US-08-072-610-1	Sequence 1, Appli
c	28	49	1.3	3337	2	US-08-719-822B-1	Sequence 1, Appli
c	29	49	1.3	3337	4	US-09-092-458-1	Sequence 1, Appli
c	30	48.8	1.3	3183	4	US-08-911-393-1	Sequence 1, Appli
	31	47.8	1.3	2824	4	US-07-757-022B-13	Sequence 13, Appl
	32	47.8	1.3	3066	4	US-07-757-022B-83	Sequence 83, Appl
	33	47.8	1.3	3117	4	US-07-757-022B-73	Sequence 73, Appl
	34	47.8	1.3	3148	4	US-07-757-022B-57	Sequence 57, Appl
	35	47.8	1.3	3420	4	US-07-757-022B-103	Sequence 103, App
	36	47.8	1.3	3813	4	US-07-757-022B-43	Sequence 43, Appl
	37	47.8	1.3	3936	4	US-07-757-022B-41	Sequence 41, Appl
	38	47.8	1.3	3942	4	US-07-757-022B-141	Sequence 141, App
	39	47.8	1.3	3945	4	US-07-757-022B-49	Sequence 49, Appl
	40	47.8	1.3	3963	4	US-07-757-022B-45	Sequence 45, Appl
	41	47.8	1.3	3963	4	US-07-757-022B-59	Sequence 59, Appl
	42	47.8	1.3	4065	4	US-07-757-022B-47	Sequence 47, Appl
	43	47.8	1.3	4086	4	US-07-757-022B-39	Sequence 39, Appl
	44	47.8	1.3	4092	4	US-07-757-022B-51	Sequence 51, Appl
	45	47.8	1.3	4215	4	US-07-757-022B-61	Sequence 61, Appl

ALIGNMENTS

```
RESULT 1
US-08-357-962-1
; Sequence 1, Application US/08357962
; Patent No. 5668263
;  GENERAL INFORMATION:
;    APPLICANT:  Hoyer, Lois
;    APPLICANT:  Livi, George
;    APPLICANT:  Shatzman, Allan
;    TITLE OF INVENTION:  CONSERVED YEAST NUCLEIC ACID SEQUENCES
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/357,962
;      FILING DATE:  16-DEC-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jervis, Herbert H
;      REGISTRATION NUMBER:  31,171
;      REFERENCE/DOCKET NUMBER:  P50278
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5019
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1071 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-357-962-1
```

```
  Query Match           28.2%;  Score 1069.4;  DB 1;  Length 1071;
  Best Local Similarity  99.9%;  Pred. No. 5.8e-252;
  Matches 1070;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1300 CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACA 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACA 60

Qy       1360 GTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACT 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACT 120

Qy       1420 GTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCT 1479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCT 180

Qy       1480 CCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACC 1539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACC 240

Qy       1540 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACT 1599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACT 300

Qy       1600 GACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCC 1659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCC 360

Qy       1660 CAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATT 1719
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATT 420

Qy       1720 ATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCC 1779
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCC 480

Qy       1780 ACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCA 1839
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCA 540

Qy       1840 CCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACT 1899
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACT 600

Qy       1900 GTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCAACT 1959
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCAACT 660

Qy       1960 GTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCTCCA 2019
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCTCCA 720

Qy       2020 CCTGGTGAAACTGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACT 2079
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        721 CCTGGTGAAACTGATACCGTTCTCATCAGAGAGCCACCAAACCATACTGTCACTACTACT 780

Qy       2080 GAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACC 2139
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACC 840

Qy       2140 GATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCT 2199
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCT 900

Qy       2200 CAATCATATGCTACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATC 2259
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CAATCATATGCTACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATC 960

Qy       2260 ATTAGAGAGCCACCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCC 2319
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATTAGAGAGCCACCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCC 1020

Qy       2320 ACAACCACCACAGTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 2370
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCACCACAGTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 1071
```

RESULT 2
US-08-878-106-1
; Sequence 1, Application US/08878106
; Patent No. 5817466
; GENERAL INFORMATION:
; APPLICANT: Hoyer, Lois
; APPLICANT: Livi, George
; APPLICANT: Shatzman, Allan
; TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,106
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/357,962
; FILING DATE: 16-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervis, Herbert H
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: P50278
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5019
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1071 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-878-106-1

```
  Query Match           28.2%;  Score 1069.4;  DB 1;  Length 1071;
  Best Local Similarity  99.9%;  Pred. No. 5.8e-252;
  Matches 1070;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1300 CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACA 1359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACA 60

Qy       1360 GTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACT 1419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACT 120

Qy       1420 GTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCT 1479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCT 180

Qy       1480 CCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACC 1539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACC 240
```

```
Qy       1540 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACT 1599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACT 300

Qy       1600 GACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCC 1659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCC 360

Qy       1660 CAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATT 1719
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATT 420

Qy       1720 ATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCC 1779
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCC 480

Qy       1780 ACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCA 1839
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCA 540

Qy       1840 CCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACT 1899
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACT 600

Qy       1900 GTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCAACT 1959
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCAACT 660

Qy       1960 GTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCTCCA 2019
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCTCCA 720

Qy       2020 CCTGGTGAAACTGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACT 2079
             ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        721 CCTGGTGAAACTGATACCGTTCTCATCAGAGAGCCACCAAACCATACTGTCACTACTACT 780

Qy       2080 GAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACC 2139
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACC 840

Qy       2140 GATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCT 2199
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCT 900

Qy       2200 CAATCATATGCTACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATC 2259
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CAATCATATGCTACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATC 960

Qy       2260 ATTAGAGAGCCACCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCC 2319
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATTAGAGAGCCACCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCC 1020

Qy       2320 ACAACCACCACAGTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 2370
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ACAACCACCACAGTTACTGCTCCTCCAGGTGGTACTGACACTGTGATTATC 1071


RESULT 3
US-08-357-962-5
; Sequence 5, Application US/08357962
; Patent No. 5668263
;  GENERAL INFORMATION:
;    APPLICANT:  Hoyer, Lois
;    APPLICANT:  Livi, George
;    APPLICANT:  Shatzman, Allan
;    TITLE OF INVENTION:  CONSERVED YEAST NUCLEIC ACID SEQUENCES
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/357,962
;      FILING DATE:  16-DEC-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jervis, Herbert H
;      REGISTRATION NUMBER:  31,171
;      REFERENCE/DOCKET NUMBER:  P50278
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5019
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  425 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-357-962-5

  Query Match             9.2%;  Score 348.2;  DB 1;  Length 425;
  Best Local Similarity   92.8%;  Pred. No. 5.5e-76;
  Matches  376;  Conservative    0;  Mismatches   28;  Indels    1;  Gaps    1;

Qy       1980 GTCTCAATCATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGT 2039
             ||| |||||||||||| |||||||| ||  | || || ||||| |||| |||||| || ||
Db         21 GTCACAATCATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGT 80

Qy       2040 TCTT-ATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCAT 2098
              |  | || ||||||||||||||||||||||||| |||||||||||||||||||| || |
Db         81 TTATCATTAGAGAGCCACCAAACCATACTGTCACCACTACTGAATACTGGTCTCAGTCCT 140

Qy       2099 ATGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACCGATACCGTTCTTATCAGAG 2158
              |||| || || || ||  |||| || ||||||||||||||||||||||||||||||||||
Db        141 ATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACCGATACCGTTCTTATCAGAG 200

Qy       2159 AGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCA 2218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCA 260

Qy       2219 CCACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGAGCCACCAAATC 2278
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        261 CCACTGTTACTGCACCACCAGGAGGTACCGATACTGTTATCATTAGAGAGCCACCAAATC 320

Qy       2279 CAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACAGTTACTG 2338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 CAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACAGTTACTG 380

Qy       2339 CTCCTCCAGGTGGTACTGACACTGTGATTATCTATGAAAGCATGT 2383
              |||||||||||||||||||||||||||||||||||||||||||||
Db        381 CTCCTCCAGGTGGTACTGACACTGTGATTATCTATGAAAGCATGT 425


RESULT 4
US-08-878-106-5
; Sequence 5, Application US/08878106
; Patent No. 5817466
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  Hoyer, Lois
;    APPLICANT:  Livi, George
;    APPLICANT:  Shatzman, Allan
;    TITLE OF INVENTION:  CONSERVED YEAST NUCLEIC ACID SEQUENCES
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406-0939
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/878,106
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/357,962
;      FILING DATE:  16-DEC-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jervis, Herbert H
;      REGISTRATION NUMBER:  31,171
;      REFERENCE/DOCKET NUMBER:  P50278
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5019
;      TELEFAX:  610-270-5090
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  425 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-878-106-5

  Query Match             9.2%;  Score 348.2;  DB 1;  Length 425;
  Best Local Similarity   92.8%;  Pred. No. 5.5e-76;
  Matches  376;  Conservative    0;  Mismatches   28;  Indels    1;  Gaps    1;

Qy       1980 GTCTCAATCATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGT 2039
              ||| |||||||||||| |||||||| ||  | || || |||| |||| |||| || ||
Db         21 GTCACAATCATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGT 80

Qy       2040 TCTT-ATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCAT 2098
              |  | || ||||||||||||||||||||||||| |||||||||||||||||||| || |
Db         81 TTATCATTAGAGAGCCACCAAACCATACTGTCACCACTACTGAATACTGGTCTCAGTCCT 140

Qy       2099 ATGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACCGATACCGTTCTTATCAGAG 2158
              |||| || || || ||  |||| || ||||||||||||||||||||||||||||||||||||
Db        141 ATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACCGATACCGTTCTTATCAGAG 200

Qy       2159 AGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCA 2218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCA 260

Qy       2219 CCACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGAGCCACCAAATC 2278
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        261 CCACTGTTACTGCACCACCAGGAGGTACCGATACTGTTATCATTAGAGAGCCACCAAATC 320

Qy       2279 CAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACAGTTACTG 2338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 CAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACCACAGTTACTG 380

Qy       2339 CTCCTCCAGGTGGTACTGACACTGTGATTATCTATGAAAGCATGT 2383
              |||||||||||||||||||||||||||||||||||||||||||||
Db        381 CTCCTCCAGGTGGTACTGACACTGTGATTATCTATGAAAGCATGT 425


RESULT 5
US-08-928-361B-2
; Sequence 2, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-1677
;      TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5511 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-928-361B-2

  Query Match             2.3%;  Score 87;  DB 3;  Length 5511;
  Best Local Similarity   46.3%;  Pred. No. 9e-12;
  Matches  285;  Conservative    0;  Mismatches  330;  Indels    0;  Gaps    0;

Qy        986 CTACAACTAGAACAGTTACAGACAGTACCACTGCTGTCACTACTTTACCATTCAATCCAA 1045
              | |||||| | ||||   |||   | ||| ||| ||  |||||||    |    |  |||
Db        974 CAACAACGACAACAACAACAACTACTACAACTACTACCACTACTACTACGACAACAACAA 1033

Qy       1046 GTGTTGATAAAACCAAAACAATCGAAATTTTGCAACCTATTCCAACCACTACCATCACAA 1105
               |  |  ||  || | |||||    || |       ||| | ||||||| || |  ||||
Db       1034 CTACTACTACGACAACAACAACAACAACTACTACTACTACTACAACCACAACAACTACAA 1093

Qy       1106 CTTCATATGTTGGTGTGACTACTTCCTATCTGACTAAGACTGCACCAATTGGTGAAACAG 1165
              |  |      |      || ||  |   |  |  ||  |  || |||       ||||
Db       1094 CCACGACAACTACAACCACAACCACAACTACCAAGAAACCAACAACAACAACAACAACAA 1153

Qy       1166 CTACTGTTATTGTTGATGTGCCATATCATACTACCACAACTGTTACCAGTGAATGGACAG 1225
              | |||  || |  |       ||      || || || |||  ||| | |      ||
Db       1154 CAACTACTACTACTACAACAACAACAACAACAACTACTACTACTACTACTACTACTA 1213

Qy       1226 GAACAATCACTACCACCACAACTCGTACCAATCCAACTGATTCAATTGACACAGTGGTGG 1285
                || |  ||||| || || ||    ||||   | ||     ||| |  ||
```

```
Db       1214 CTACTACTACTACTACTACCACAACAACCACAACCACAACCACAACTACCAAGAAACCAA 1273

Qy       1286 TACAAGTTCCACTGCCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTG 1345

Db       1274 CAACAACAACAACAACAACAACAACAACTACTACTACAACCACGACAACAACAACCACGA 1333

Qy       1346 CTACAACCACTACAGTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGC 1405

Db       1334 CAACCACAACCACAACTACAACTACCAAGAAACCAACAACTACTACTACTACTACCACAA 1393

Qy       1406 CACCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTA 1465

Db       1394 CAACAACAACTACTACTACTACCACAACAACAACAACTACTACTACTACTACAACCACAA 1453

Qy       1466 CTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAA 1525

Db       1454 CCACAACCACAACCGCAACCACAACTACCAAGAAACCAACAACAACAACAACAACTACTA 1513

Qy       1526 CTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTC 1585

Db       1514 CTACTACTACAACCAAGAAACCAACAACAACTACCACTGCCACAACAACAACTACTACTT 1573

Qy       1586 CTCCAGGTGGTACTG 1600

Db       1574 CTGAAACTGAGAGTG 1588


RESULT 6
US-08-928-361B-1
; Sequence 1, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-1677
;      TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7334 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-928-361B-1

  Query Match             2.3%;  Score 87;  DB 3;  Length 7334;
  Best Local Similarity   46.3%;  Pred. No. 1e-11;
  Matches  285;  Conservative    0;  Mismatches  330;  Indels    0;  Gaps    0;

Qy        986 CTACAACTAGAACAGTTACAGACAGTACCACTGCTGTCACTACTTTACCATTCAATCCAA 1045

Db       2642 CAACAACGACAACAACAACAACTACTACAACTACTACCACTACTACTACGACAACAACAA 2701

Qy       1046 GTGTTGATAAAACCAAAACAATCGAAATTTTGCAACCTATTCCAACCACTACCATCACAA 1105

Db       2702 CTACTACTACGACAACAACAACAACAACTACTACTACTACTACAACCACAACAACTACAA 2761

Qy       1106 CTTCATATGTTGGTGTGACTACTTCCTATCTGACTAAGACTGCACCAATTGGTGAAACAG 1165

Db       2762 CCACGACAACTACAACCACAACCACAACTACCAAGAAACCAACAACAACAACAACAACAA 2821

Qy       1166 CTACTGTTATTGTTGATGTGCCATATCATACTACCACAACTGTTACCAGTGAATGGACAG 1225

Db       2822 CAACTACTACTACTACAACAACAACAACAACAACTACTACTACTACTACTACTACTACTA 2881

Qy       1226 GAACAATCACTACCACCACAACTCGTACCAATCCAACTGATTCAATTGACACAGTGGTGG 1285

Db       2882 CTACTACTACTACTACTACCACAACAACCACAACCACAACCACAACTACCAAGAAACCAA 2941

Qy       1286 TACAAGTTCCACTGCCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTG 1345

Db       2942 CAACAACAACAACAACAACAACAACAACTACTACTACAACCACGACAACAACAACCACGA 3001

Qy       1346 CTACAACCACTACAGTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGC 1405

Db       3002 CAACCACAACCACAACTACAACTACCAAGAAACCAACAACTACTACTACTACTACCACAA 3061

Qy       1406 CACCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTA 1465

Db       3062 CAACAACAACTACTACTACTACCACAACAACAACAACTACTACTACTACTACAACCACAA 3121

Qy       1466 CTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAA 1525

Db       3122 CCACAACCACAACCGCAACCACAACTACCAAGAAACCAACAACAACAACAACAACTACTA 3181

Qy       1526 CTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTC 1585

Db       3182 CTACTACTACAACCAAGAAACCAACAACAACTACCACTGCCACAACAACAACTACTACTT 3241

Qy       1586 CTCCAGGTGGTACTG 1600

Db       3242 CTGAAACTGAGAGTG 3256


RESULT 7
US-08-700-651-1
; Sequence 1, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
;  APPLICANT: PETERSEN, CAROLYN
;  APPLICANT: LEECH, JAMES
;  APPLICANT: NELSON, RICHARD, C.
;  APPLICANT: GUT, JIRI
;  TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
;  TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
;  TITLE OF INVENTION: INFECTIONS
;  FILE REFERENCE: 480.19-4(HV)
;  CURRENT APPLICATION NUMBER: US/08/700,651B
;  CURRENT FILING DATE: 1997-08-14
;  EARLIER APPLICATION NUMBER: 08/415,751
;  EARLIER FILING DATE: 1995-04-03
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 5163
;   TYPE: DNA
;   ORGANISM: Cryptosporidium parvum
US-08-700-651-1
```

```
  Query Match             2.3%;  Score 86.2;  DB 3;  Length 5163;
  Best Local Similarity   47.3%;  Pred. No. 1.4e-11;
  Matches  311;  Conservative    0;  Mismatches  328;  Indels   18;  Gaps    1;

Qy       1406 CACCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTA 1465
Db        588 CAACAACTACTACTACTACTACTACGACAACAACAACAACAACGACAACAACAACAACTA 647

Qy       1466 CTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAA 1525
Db        648 CTACAACTACTACCACTACTACTACGACAACAACAACAACAACAACAACAACAACAACAA 707

Qy       1526 CTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTC 1585
Db        708 CAACAACAACAACAACAACAACAACAACGACTACTACTACTACTACTACTACTACTACTA 767

Qy       1586 CTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCA 1645
Db        768 CTACTACTACTACAACCACAACAACTACAACCACAACTACAACCACAACAACTACAACCA 827

Qy       1646 CTGAATATTGGTCCCAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCA 1705
Db        828 CAACAACAACAACCACAACAACTACAACCACAACTACAACCACAACAACTACAACCACAA 887

Qy       1706 CTGACTCAGTAATTATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGT 1765
Db        888 CAACCACAACCACAACCACAACTACCAAGAAACCAACAACAACAACTACAACAACAACAA 947

Qy       1766 CACAATCATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTA 1825
Db        948 CAACAACAACAACTACTACTACAACCACCACAACAACAACAACAACAACTACAACTA 1007

Qy       1826 TCATTAGAGAGCCACCAAACCACACT----------------GTCACTACTACTGAGT 1867
Db       1008 CCAAGAAACCAACAACTACTACTACTACCACAACAACAACAACTACTACTACTACCA 1067

Qy       1868 ATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGCACTGATA 1927
Db       1068 CAACAACAACAACTACTACTACTACTACAACAACAACAACAACAACAACAACAACAACAA 1127

Qy       1928 CTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACTACTGAATACTGGTCTCAAT 1987
Db       1128 CTACCACGAAACCAACAACAACAACAACAACTACTACTACTACTACAACCAAGAAACCAA 1187

Qy       1988 CATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGTTCTTA 2044
Db       1188 CAACAACTACCACTGCCACAACAACAACTACTACTTCTGAAACTGAGAGTGTAATTA 1244


RESULT 8
US-08-928-361B-4
; Sequence 4, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-1677
;      TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5163 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-928-361B-4

  Query Match             2.3%;  Score 86.2;  DB 3;  Length 5163;
  Best Local Similarity   47.3%;  Pred. No. 1.4e-11;
  Matches  311;  Conservative    0;  Mismatches  328;  Indels   18;  Gaps    1;

Qy       1406 CACCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTA 1465
Db        587 CAACAACTACTACTACTACTACTACGACAACAACAACAACAACGACAACAACAACAACTA 646

Qy       1466 CTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAA 1525
Db        647 CTACAACTACTACCACTACTACTACGACAACAACAACAACAACAACAACAACAACAACAA 706

Qy       1526 CTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTC 1585
Db        707 CAACAACAACAACAACAACAACAACAACGACTACTACTACTACTACTACTACTACTACTA 766

Qy       1586 CTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCA 1645
Db        767 CTACTACTACTACAACCACAACAACTACAACCACAACTACAACCACAACAACTACAACCA 826

Qy       1646 CTGAATATTGGTCCCAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCA 1705
Db        827 CAACAACAACAACCACAACAACTACAACCACAACTACAACCACAACAACTACAACCACAA 886

Qy       1706 CTGACTCAGTAATTATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGT 1765
Db        887 CAACCACAACCACAACCACAACTACCAAGAAACCAACAACAACAACTACAACAACAACAA 946

Qy       1766 CACAATCATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTA 1825
Db        947 CAACAACAACAACTACTACTACAACCACCACAACAACAACAACAACAACAACTACAACTA 1006

Qy       1826 TCATTAGAGAGCCACCAAACCACACT----------------GTCACTACTACTGAGT 1867
Db       1007 CCAAGAAACCAACAACTACTACTACTACCACAACAACAACAACTACTACTACTACCA 1066

Qy       1868 ATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGCACTGATA 1927
Db       1067 CAACAACAACAACTACTACTACTACTACAACAACAACAACAACAACAACAACAACAACAA 1126

Qy       1928 CTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACTACTGAATACTGGTCTCAAT 1987
Db       1127 CTACCACGAAACCAACAACAACAACAACAACTACTACTACTACTACAACCAAGAAACCAA 1186

Qy       1988 CATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGTTCTTA 2044
Db       1187 CAACAACTACCACTGCCACAACAACAACTACTACTTCTGAAACTGAGAGTGTAATTA 1243


RESULT 9
```

```
US-08-700-651-2
; Sequence 2, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
;  APPLICANT: PETERSEN, CAROLYN
;  APPLICANT: LEECH, JAMES
;  APPLICANT: NELSON, RICHARD, C.
;  APPLICANT: GUT, JIRI
;  TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
;  TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
;  TITLE OF INVENTION: INFECTIONS
;  FILE REFERENCE: 480.19-4(HV)
;  CURRENT APPLICATION NUMBER: US/08/700,651B
;  CURRENT FILING DATE: 1997-08-14
;  EARLIER APPLICATION NUMBER: 08/415,751
;  EARLIER FILING DATE: 1995-04-03
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 5318
;   TYPE: DNA
;   ORGANISM: Cryptosporidium parvum
US-08-700-651-2

  Query Match             2.3%;  Score 86.2;  DB 3;  Length 5318;
  Best Local Similarity   47.3%;  Pred. No. 1.4e-11;
  Matches  311;  Conservative    0;  Mismatches  328;  Indels   18;  Gaps    1;

Qy       1406 CACCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTA 1465
             || |||    ||||   ||||||||   |        ||  | |     | || || ||||
Db        588 CAACAACTACTACTACTACTACGACAACAACAACAACAACGACAACAACAACAACTA 647

Qy       1466 CTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAA 1525
             ||   ||| || |  |   |  |||     ||  ||  |  | |  | || |||   |||
Db        648 CTACAACTACTACCACTACTACTACGACAACAACAACAACAACAACAACAACAACAACAA 707

Qy       1526 CTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTC 1585
             |    || |||||           |      |   | ||||||||||||||   |||| ||
Db        708 CAACAACAACAACAACAACAACAACAACGACTACTACTACTACTACTACTACTACTACTA 767

Qy       1586 CTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCA 1645
             || |   |  |||   | ||  || ||  |    | || ||| | ||||    || ||||
Db        768 CTACTACTACTACAACCACAACAACTACAACCACAACTACAACCACAACAACTACAACCA 827

Qy       1646 CTGAATATTGGTCCCAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCA 1705
             |   |        ||  | |     ||||||||||||||    ||  |  || ||       |
Db        828 CAACAACAACAACCACAACAACTACAACCACAACTACAACCACAACAACTACAACCACAA 887

Qy       1706 CTGACTCAGTAATTATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGT 1765
             |   | ||   |  | || |     |    || |  ||    || |||||   |
Db        888 CAACCACAACCACAACCACAACTACCAAGAAACCAACAACAACAACTACAACAACAACAA 947

Qy       1766 CACAATCATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTA 1825
             ||  | ||    | || ||||| ||    ||  || || ||       ||    ||  ||
Db        948 CAACAACAACAACTACTACTACAACCACCACAACAACAACAACAACAACAACTACAACTA 1007

Qy       1826 TCATTAGAGAGCCACCAAACCACACT----------------GTCACTACTACTGAGT 1867
              ||  | |    || | |     |||                |||||||||
Db       1008 CCAAGAAACCAACAACTACTACTACTACTACCACAACAACAACAACTACTACTACTACCA 1067

Qy       1868 ATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGCACTGATA 1927
                |  |   | ||||||| |||||  |||    || |||     ||    |
Db       1068 CAACAACAACAACTACTACTACTACTACAACAACAACAACAACAACAACAACAACAACAA 1127

Qy       1928 CTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACTACTGAATACTGGTCTCAAT 1987
              ||   |  |        | || |||   |||||   || ||||||  |  |  |   | |
Db       1128 CTACCACGAAACCAACAACAACAACAACAACTACTACTACTACTACAACCAAGAAACCAA 1187

Qy       1988 CATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGTTCTTA 2044
              ||    | |||||| | || |  ||  || |  ||  ||||||||| |  ||  |||
Db       1188 CAACAACTACCACTGCCACAACAACAACTACTACTTCTGAAACTGAGAGTGTAATTA 1244
```

```
RESULT 10
US-08-928-361B-3
; Sequence 3, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-1677
;      TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5318 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-928-361B-3

  Query Match             2.3%;  Score 86.2;  DB 3;  Length 5318;
  Best Local Similarity   47.3%;  Pred. No. 1.4e-11;
  Matches  311;  Conservative    0;  Mismatches  328;  Indels   18;  Gaps    1;

Qy       1406 CACCAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTA 1465
             || |||    ||||   ||||||||   |        ||  | |     | || || ||||
Db        587 CAACAACTACTACTACTACTACGACAACAACAACAACAACGACAACAACAACAACTA 646

Qy       1466 CTGTTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAA 1525
             ||   ||| || |  |   |  |||     ||  ||  |  | |  | || |||   |||
Db        647 CTACAACTACTACCACTACTACTACGACAACAACAACAACAACAACAACAACAACAACAA 706

Qy       1526 CTGTCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTC 1585
             |    || |||||           |      |   | ||||||||||||||   |||| ||
Db        707 CAACAACAACAACAACAACAACAACAACGACTACTACTACTACTACTACTACTACTACTA 766

Qy       1586 CTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCA 1645
             || |   |  |||   | ||  || ||  |    | || ||| | ||||    || ||||
Db        767 CTACTACTACTACAACCACAACAACTACAACCACAACTACAACCACAACAACTACAACCA 826

Qy       1646 CTGAATATTGGTCCCAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCA 1705
             |   |        ||  | |     ||||||||||||||    ||  |  || ||       |
Db        827 CAACAACAACAACCACAACAACTACAACCACAACTACAACCACAACAACTACAACCACAA 886

Qy       1706 CTGACTCAGTAATTATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGT 1765
             |   | ||   |  | || |     |    || |  ||    || |||||   |
Db        887 CAACCACAACCACAACCACAACTACCAAGAAACCAACAACAACAACTACAACAACAACAA 946
```

```
Qy       1766 CACAATCATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTA 1825
              ||  | ||    | || |||||  ||    ||  || || ||      ||    ||   ||
Db        947 CAACAACAACAACTACTACTACAACCACCACAACAACAACAACAACAACAACTACAACTA 1006

Qy       1826 TCATTAGAGAGCCACCAAACCACACT-----------------GTCACTACTACTGAGT 1867
               ||  | |    || | |     |||                 |||||||||
Db       1007 CCAAGAAACCAACAACTACTACTACTACTACCACAACAACAACAACTACTACTACTACCA 1066

Qy       1868 ATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGCACTGATA 1927
                 |  | |   | |||||| |||||   |||     || |||     ||    |
Db       1067 CAACAACAACAACTACTACTACTACTACAACAACAACAACAACAACAACAACAACAACAA 1126

Qy       1928 CTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACTACTGAATACTGGTCTCAAT 1987
              ||   |  |       | || |||   |||||   || ||||||  |  |  |   | |
Db       1127 CTACCACGAAACCAACAACAACAACAACAACTACTACTACTACTACAACCAAGAAACCAA 1186

Qy       1988 CATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAAACTGATACCGTTCTTA 2044
              ||    | |||||| | || |  ||  || |  ||  |||||||| |  ||  |||
Db       1187 CAACAACTACCACTGCCACAACAACAACTACTACTTCTGAAACTGAGAGTGTAATTA 1243
```

RESULT 11
US-08-357-962-2
; Sequence 2, Application US/08357962
; Patent No. 5668263
; GENERAL INFORMATION:
; APPLICANT: Hoyer, Lois
; APPLICANT: Livi, George
; APPLICANT: Shatzman, Allan
; TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/357,962
; FILING DATE: 16-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervis, Herbert H
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: P50278
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5019
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-357-962-2

Query Match 2.3%; Score 85.2; DB 1; Length 108;
Best Local Similarity 60.2%; Pred. No. 5.4e-12;
Matches 65; Conservative 41; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1300 CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACA 1359
              |||||:|::||:||:|::||:||:||:||:||||| ||:|| |::||:||:||:||:||:
Db          1 CCAAAYCMHACWGTYASYACHACYGARTAYTGGTCNCARTCNTWYGCHACHACHACHACH 60

Qy       1360 GTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCA 1407
              :|:||:|::||:||::|::|:||:||::|:||::|:||:||:||:||:
Db         61 RTDACYGSWCCWCCWRGWRGHACYGAYWCHGTDMTYATYAGRGARCCW 108
```

RESULT 12
US-08-878-106-2
; Sequence 2, Application US/08878106
; Patent No. 5817466
; GENERAL INFORMATION:
; APPLICANT: Hoyer, Lois
; APPLICANT: Livi, George
; APPLICANT: Shatzman, Allan
; TITLE OF INVENTION: CONSERVED YEAST NUCLEIC ACID SEQUENCES
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,106
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/357,962
; FILING DATE: 16-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jervis, Herbert H
; REGISTRATION NUMBER: 31,171
; REFERENCE/DOCKET NUMBER: P50278
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5019
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 108 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
US-08-878-106-2

Query Match 2.3%; Score 85.2; DB 1; Length 108;
Best Local Similarity 60.2%; Pred. No. 5.4e-12;
Matches 65; Conservative 41; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1300 CCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACA 1359
              |||||:|::||:||:|::||:||:||:||:||||| ||:|| |::||:||:||:||:||:
Db          1 CCAAAYCMHACWGTYASYACHACYGARTAYTGGTCNCARTCNTWYGCHACHACHACHACH 60

Qy       1360 GTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCA 1407
              :|:||:|::||:||::|::|:||:||::|:||::|:||:||:||:||:
Db         61 RTDACYGSWCCWCCWRGWRGHACYGAYWCHGTDMTYATYAGRGARCCW 108
```

```
RESULT 13
US-08-145-705A-32/c
; Sequence 32, Application US/08145705A
; Patent No. 5489513
;  GENERAL INFORMATION:
;    APPLICANT:  Springer, Wolfgang; Plempel,Manfred;
;    APPLICANT:  L bberding, Antonius
;    TITLE OF INVENTION:  SPECIFIC GENE PROBES AND
;    TITLE OF INVENTION:  PROCESSES FOR THE DIAGNOSTIC
;    TITLE OF INVENTION:  INVESTIGATION OF CANDIDA
;    TITLE OF INVENTION:  ALBICANS
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPRUNG HORN KRAMER & WOODS
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.4 MB storage
;      COMPUTER:  NEC PowerMate 1 Plus
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/145,705A
;      FILING DATE:  October 28, 1993
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  German P 42 36 708.5
;      FILING DATE:  October 30, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  Bayer 8885-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  32:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  100 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Candida albicans
US-08-145-705A-32

  Query Match             2.2%;  Score 82.4;  DB 1;  Length 100;
  Best Local Similarity   89.0%;  Pred. No. 2.5e-11;
  Matches   89;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       1170 TGTTATTGTTGATGTGCCATATCATACTACCACAACTGTTACCAGTGAATGGACAGGAAC 1229
             ||||||||||||||| |||||||| |||||||||||||||| |||||||||||||||||
Db        100 TGTTATTGTTGATGTTCCATATCACACTACCACAACTGTTACTAGTGAATGGACAGGAAC 41

Qy       1230 AATCACTACCACCACAACTCGTACCAATCCAACTGATTCA 1269
              ||||||||| || ||||| |  || |||||||| | ||||
Db         40 AATCACTACTACTACAACACAAACTAATCCAACAGGTTCA 1


RESULT 14
US-08-145-705A-34/c
; Sequence 34, Application US/08145705A
; Patent No. 5489513
;  GENERAL INFORMATION:
;    APPLICANT:  Springer, Wolfgang; Plempel,Manfred;
;    APPLICANT:  L bberding, Antonius
;    TITLE OF INVENTION:  SPECIFIC GENE PROBES AND
;    TITLE OF INVENTION:  PROCESSES FOR THE DIAGNOSTIC
;    TITLE OF INVENTION:  INVESTIGATION OF CANDIDA
;    TITLE OF INVENTION:  ALBICANS
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPRUNG HORN KRAMER & WOODS
;      STREET:  660 White Plains Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10591-5144
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.4 MB storage
;      COMPUTER:  NEC PowerMate 1 Plus
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/145,705A
;      FILING DATE:  October 28, 1993
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  German P 42 36 708.5
;      FILING DATE:  October 30, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kurt G. Briscoe
;      REGISTRATION NUMBER:  33,141
;      REFERENCE/DOCKET NUMBER:  Bayer 8885-KGB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 332-1700
;      TELEFAX:  (914) 332-1844
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  100 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Candida albicans
US-08-145-705A-34

  Query Match             2.2%;  Score 82.4;  DB 1;  Length 100;
  Best Local Similarity   89.0%;  Pred. No. 2.5e-11;
  Matches   89;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy        970 GGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCTGTCACTACT 1029
              |||||||||||||| |||||||||||||||||||||| ||||| |||||||| ||||||
Db        100 GGTATTGTCATTGTGGCTACAACTAGAACAGTTACAGATAGTACTACTGCTGTGACTACT 41

Qy       1030 TTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCG 1069
              |||||||||||| |   |||||| ||||| |||||||| |
Db         40 TTACCATTCAATTCCGATGTTGACAAAACTAAAACAATTG 1


RESULT 15
US-08-145-705A-33/c
; Sequence 33, Application US/08145705A
; Patent No. 5489513
;  GENERAL INFORMATION:
;    APPLICANT:  Springer, Wolfgang; Plempel,Manfred;
;    APPLICANT:  L bberding, Antonius
;    TITLE OF INVENTION:  SPECIFIC GENE PROBES AND
;    TITLE OF INVENTION:  PROCESSES FOR THE DIAGNOSTIC
;    TITLE OF INVENTION:  INVESTIGATION OF CANDIDA
;    TITLE OF INVENTION:  ALBICANS
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SPRUNG HORN KRAMER & WOODS
```

; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 MB storage
; COMPUTER: NEC PowerMate 1 Plus
; OPERATING SYSTEM: DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/145,705A
; FILING DATE: October 28, 1993
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: German P 42 36 708.5
; FILING DATE: October 30, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: Bayer 8885-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; TELEX:
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Candida albicans
US-08-145-705A-33

Query Match 2.0%; Score 76; DB 1; Length 100;
Best Local Similarity 85.0%; Pred. No. 9.3e-10;
Matches 85; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
Qy       1070 AAATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTT 1129
              |||||||||||||| |||||||| |||||||||||||| ||||||||||||||||||||
Db        100 AAATTTTGCAACCCATTCCAACAACTACCATCACAACATCATATGTTGGTGTGACTACTT 41

Qy       1130 CCTATCTGACTAAGACTGCACCAATTGGTGAAACAGCTAC 1169
              ||||    ||  | |||| |||||| ||  |||| |||||
Db         40 CCTACAGAACCCAAACTGTACCAATAGGACAAACTGCTAC 1
```

Search completed: June 20, 2003, 06:16:05
Job time : 134 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2003, 05:18:05 ; Search time 588 Seconds
(without alignments)
9448.434 Million cell updates/sec

Title: US-09-715-876-7
Perfect score: 3786
Sequence: 1 atgcttcaacaatttacatt.....tgtccttgttcatttagtga 3786

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	78.4	2.1	5403	10	US-09-745-008-33	Sequence 33, Appl
2	74.4	2.0	684973	10	US-09-263-959-1	Sequence 1, Appli
c 3	62.2	1.6	2014	10	US-09-842-552-22	Sequence 22, Appl
4	55.4	1.5	436	10	US-09-864-761-2885	Sequence 2885, Ap
5	54.6	1.4	867	10	US-09-216-393-340	Sequence 340, App
6	54.6	1.4	867	10	US-09-216-393-342	Sequence 342, App
7	54.6	1.4	1397	10	US-09-216-393-343	Sequence 343, App
c 8	54.6	1.4	1397	10	US-09-216-393-345	Sequence 345, App
9	52.8	1.4	4104	10	US-09-801-368-107	Sequence 107, App
c 10	52.2	1.4	1059	12	US-10-073-256-54	Sequence 54, Appl
c 11	52	1.4	2015	10	US-09-842-552-79	Sequence 79, Appl
12	50.8	1.3	15720	9	US-10-025-380-1058	Sequence 1058, Ap
13	50.8	1.3	15720	10	US-09-922-217-1058	Sequence 1058, Ap
14	50.8	1.3	15720	10	US-09-833-263-1058	Sequence 1058, Ap
15	50.6	1.3	1236	9	US-10-077-584-3	Sequence 3, Appli
16	50.2	1.3	4197	10	US-09-137-531-7	Sequence 7, Appli
17	50.2	1.3	4197	10	US-09-137-531-8	Sequence 8, Appli
c 18	48.8	1.3	3183	10	US-09-955-909-1	Sequence 1, Appli
c 19	48.6	1.3	2614	9	US-09-822-846-491	Sequence 491, App
20	48.2	1.3	1101	10	US-09-874-062-2	Sequence 2, Appli
c 21	48.2	1.3	6604	10	US-09-880-107-1748	Sequence 1748, Ap
c 22	48	1.3	3331	10	US-09-864-761-19481	Sequence 19481, A
23	48	1.3	7104	10	US-09-815-242-4580	Sequence 4580, Ap
24	48	1.3	7107	10	US-09-815-242-8291	Sequence 8291, Ap
25	47.8	1.3	460	10	US-09-864-761-19383	Sequence 19383, A
26	47.8	1.3	2824	12	US-10-124-557-13	Sequence 13, Appl
27	47.8	1.3	3066	12	US-10-124-557-83	Sequence 83, Appl
28	47.8	1.3	3117	12	US-10-124-557-73	Sequence 73, Appl
29	47.8	1.3	3148	12	US-10-124-557-57	Sequence 57, Appl
30	47.8	1.3	3420	12	US-10-124-557-103	Sequence 103, App
31	47.8	1.3	3813	12	US-10-124-557-43	Sequence 43, Appl
32	47.8	1.3	3936	12	US-10-124-557-41	Sequence 41, Appl
33	47.8	1.3	3942	12	US-10-124-557-141	Sequence 141, App
34	47.8	1.3	3945	12	US-10-124-557-49	Sequence 49, Appl
35	47.8	1.3	3963	12	US-10-124-557-45	Sequence 45, Appl
36	47.8	1.3	3963	12	US-10-124-557-59	Sequence 59, Appl
37	47.8	1.3	4065	12	US-10-124-557-47	Sequence 47, Appl
38	47.8	1.3	4086	12	US-10-124-557-39	Sequence 39, Appl
39	47.8	1.3	4092	12	US-10-124-557-51	Sequence 51, Appl
40	47.8	1.3	4215	12	US-10-124-557-61	Sequence 61, Appl
41	47.8	1.3	4575	12	US-10-044-090-303	Sequence 303, App
42	47.8	1.3	5008	12	US-10-124-557-1	Sequence 1, Appli
c 43	47	1.2	5361	9	US-09-742-096-2	Sequence 2, Appli
c 44	47	1.2	6060	7	US-08-781-986A-534	Sequence 534, App
c 45	47	1.2	6152	9	US-09-742-096-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-745-008-33/c
; Sequence 33, Application US/09745008
; Patent No. US20020137667A1
; GENERAL INFORMATION:
; APPLICANT: Chuenkova, Marina
; APPLICANT: Pereira, Miercio A.
; TITLE OF INVENTION: T. Cruzi-Derived Neurotrophic Agents and
; TITLE OF INVENTION: Methods of Use Therefor
; FILE REFERENCE: 1322.1028-001
; CURRENT APPLICATION NUMBER: US/09/745,008
; CURRENT FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/172,881
; PRIOR FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 5403
; TYPE: DNA
; ORGANISM: Trypanosoma cruzi
US-09-745-008-33

Query Match 2.1%; Score 78.4; DB 10; Length 5403;
Best Local Similarity 42.2%; Pred. No. 3.8e-07;
Matches 442; Conservative 0; Mismatches 606; Indels 0; Gaps 0;

```
Qy       1309 ACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTACTGCT 1368
              || ||  | ||| |||      ||  ||       | ||     |||| | || |  |
Db       3789 ACCGTGGGCACTGCTGTCAGCGGGAGTCGAGGGCGTACTGTGGGCACTGCTGTCAGCGGG 3730

Qy       1369 CCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCACTACT 1428
                || |||  ||||||          | |||| | |   | |   | ||||||    |||
Db       3729 AGTCGAGGGCGTACCGTGGGCACTGCTGTCAGCGGGAGTCGAGGGCGTACTGTGGGCACT 3670

Qy       1429 ACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGGTGGT 1488
               |||   |  ||   | |     | ||      ||| |||| |  |    || |||  ||
Db       3669 GCTGTCAACGGGAGTCGAGGGCGTACTGTGGGCACTGCTGTCAGCGGGAGTCGAGGGCGT 3610

Qy       1489 ACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTATTGG 1548
              || |          | |||| |       |   |  |||||    ||  | |   |  ||
Db       3609 ACCGTGGGCACTGCTGTCAGCGGGAGTTGAGGGCGTACTGTGGGCACTGCTGTCAACGGG 3550
```

```
Qy       1549 TCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTCAGTA 1608
Db       3549 AGTTGAGGGCGTACTGTGGGCACTGCTGTCAGCGGGAGTCGAGGGCGTACTGTGGGCACT 3490

Qy       1609 ATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAATCTTAC 1668
Db       3489 GCTGTCAGCGGGAGTTGAGGGCGTACTGTGGGCACTGCTGTCAGCGGGAGTTGAGGGCGT 3430

Qy       1669 GCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATCAGAGAA 1728
Db       3429 ACTGTGGGCACTGCTGTCAACGGGAGTTGAGGGCGTACTGTGGGCACTGCTGTCAGCGGG 3370

Qy       1729 CCACCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACCACTACC 1788
Db       3369 AGTCGAGGGCGTACTGTGGGCACTGCTGTCAGCGGGAGTCGAGGGCGTACTGTGGGCACT 3310

Qy       1789 ACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCACCAAACCAC 1848
Db       3309 GCTGTCAGCGGGAGTCGAGGGCGTACCGTGGGCACTGCTGTCAACGGGAGTTGAGGGCGT 3250

Qy       1849 ACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACTGTAACTGGT 1908
Db       3249 ACCGTGGGCACTGCTGTCAGCGGGAGTTGAGGGCGTACTGTGGGCACTGCTGTCAACGGG 3190

Qy       1909 CCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCAACTGTCACCACT 1968
Db       3189 AGTTGAGGGCGTACTGTGGGCACTGCTGTCAGCGGGAGTCGAGGGCGTACTGTGGGCACT 3130

Qy       1969 ACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCTCCACCTGGTGAA 2028
Db       3129 GCTGTCAACGGGAGTTGAGGGCGTACCGTGGGCACTGCTGTCAGCGGGAGTTGAGGGCGT 3070

Qy       2029 ACTGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGG 2088
Db       3069 ACTGTGGGCACTGCTGTCAGCGGGAGTCGAGGGCGTACTGTGGGCACTGCTGTCAACGGG 3010

Qy       2089 TCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAAACCGATACCGTT 2148
Db       3009 AGTTGAGGGCGTACCGTGGGCACTGCTGTCAACGGGAGTCGAGGGCGTACTGTGGGCACT 2950

Qy       2149 CTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGGTCTCAATCATAT 2208
Db       2949 GCTGTCAACGGGAGTTGAGGGCGTACCGTGGGCACTGCTGTCAACGGGAGTCGAGGGCGT 2890

Qy       2209 GCTACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTTATCATTAGAGAG 2268
Db       2889 ACTGTGGGCACTGCTGTCAACGGGAGTTGAGGGCGTACCGTGGGCACTGCTGTCAGCGGG 2830

Qy       2269 CCACCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTTGCCACAACCACC 2328
Db       2829 AGTCGAGGGCGTACTGTGGGCACTGCTGTCAGCGGGAGTCGAGGGCGTACCGTGGGCACT 2770

Qy       2329 ACAGTTACTGCTCCTCCAGGTGGTACTG 2356
Db       2769 GCTGTCAACGGGAGTCGAGGGCGTACTG 2742
```

```
RESULT 2
US-09-263-959-1
; Sequence 1, Application US/09263959
; Patent No. US20020150891A1
;  GENERAL INFORMATION:
;    APPLICANT:  Hood, Leroy E.
;    APPLICANT:  Rowen, Lee
;    APPLICANT:  Koop, Ben F.
;    TITLE OF INVENTION:  DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH UTI
;    NUMBER OF SEQUENCES:  1279
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Seed and Berry LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  US
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/263,959
;      FILING DATE:  05-MAR-1999
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McMasters, David D.
;      REGISTRATION NUMBER:  33,963
;      REFERENCE/DOCKET NUMBER:  920010.426C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  684973 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-263-959-1

  Query Match             2.0%;  Score 74.4;  DB 10;  Length 684973;
  Best Local Similarity   43.2%;  Pred. No. 6e-05;
  Matches  626;  Conservative    0;  Mismatches  801;  Indels   21;  Gaps    5;

Qy        755 ACACTAAAACTTGTACATCTAATGGAATTCAGATTAAATATCAAAATGTACCTGCTGGTT 814
Db      12574 ATACTACCACTGGCACTACTGATACTGTTCCTATCACAACCACATCTTTCCCAAGTACTA 12633

Qy        815 ATCGTCCATTTATTGATGCTTATATTTCTGCTACAGATGTTAACCAATATACTTTAGCAT 874
Db      12634 CTAGTGTTACAACTAATACTACTGTTCCTGATACAACTTCT---CCTTTCCCTACAAGTA 12690

Qy        875 ATACCAATGATTATACTTGTGCTGGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGAT 934
Db      12691 CTACTAATGCTAGCACTAATGCTACTGTTCCTATCACAACCACACCTTTCCCAACAAGTA 12750

Qy        935 GGACTGGATACAAGAATAGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTA 994
Db      12751 CTATTGGTGTTACAACTAATGCTACTGTTCCCAATACAACTGCCCCTTTCCCAACAAATG 12810

Qy        995 GAACAGTTACAGACAGTACCACTGCTG------TCACTACTTTACCATTCAATCCAAGTG 1048
Db      12811 CTAGTACTGCTAGCACTAATGCTACTGTTCCTATCACAACCACATGTTTTGCAACAAGTA 12870

Qy       1049 TTGATAAAACCAAAACAATCGAAATTTTGCAACCTATTCCAACCACTACCATCACAACTT 1108
Db      12871 CTATTGGTGTTACAACTAATGCTACTGTTCCCGATACAACTGCCCCTTTCCCAACAAATA 12930

Qy       1109 CATATGTTGGTGTGACTACTTCCTATCTGACTAAGACTGCACCAATTGGTGAAACAGCTA 1168
Db      12931 CTACTACTGCTAGCACTAATGCTACTATTCCTATCACAACCACACCTTTTGCAACAAGTA 12990

Qy       1169 CTGTTATTGTTGATGTGCCATATCATACTACCACAACTGTTACCAGTGAATGGACAGGAA 1228
Db      12991 CTATTAGTGTTACAACTAGTACTACTGTTCCTGATACAACTGCTCCTTTCCCTACAAGTA 13050

Qy       1229 CAATCACTACCACCACAACTCGTACCAATCCAACTGATTCAATTGACACAGTGGTGGTAC 1288
Db      13051 CTACTAGTGCTAGCACTAATGCTACCCCTGTTCCTATCACAA---CCACACTTTTTGCAA 13107

Qy       1289 AAGTTCCACTGCCAAATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTA 1348
Db      13108 CAAGTACTATTGGTGTTACAACTGGTACTACTGTTCCTGATACAACTGCTCCTTTCCCTA 13167

Qy       1349 CAACCACTACAGTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCAC 1408
Db      13168 CAAGTACTAC---TAGTACTAGCACTAGTGCTACTGTTCCTATTACAACCACACCTTCCC 13224
```

```
Qy       1409 CAAACCATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTG 1468
             | |   ||||||    | |||        |     |      |   | | ||||   || ||
Db      13225 CTACAAATACTGCTGATGCTAACACTAGTAATACTGTTCCTAATACCACTATGCCTTCTC 13284

Qy       1469 TTACTGCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTG 1528
              |||   |  | |   ||  | |       |  |   |         ||   | |||  ||
Db      13285 CTACAAGTAGTACTACTGTGAGTACTATTGCTACCGTTCCCATTTCAGTGACTCCTTCTC 13344

Qy       1529 TCACTACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTC 1588
              | ||  |  ||  |   ||    |        | ||     || | |   ||||  || |||
Db      13345 TGACAAGTACTGCTGATGCCACCATTAGTACTACTGTACTTATTGCCACTACTTCTTCTC 13404

Qy       1589 CAGGTGGTACTGACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACTG 1648
              |  ||||||      ||  |   |  ||     ||   ||     |     ||  |   |
Db      13405 TAACAGGTACTACTGATGTTAGCACTAGTACTACTATTAATAATATAAGTACTCCTGTTC 13464

Qy       1649 AATATTGGTCCCAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTG 1708
              ||       |      |      ||  |  |||| |||  ||  |    | |    | ||
Db      13465 AAACAAATACTACTAATGCTAGCACTAGTACTAATGTTGCTAATATAACTGCTACCTCTC 13524

Qy       1709 ACTCAGTAATTATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGTCAC 1768
              |  ||   | |              || ||  | |||||  |   |||  |        || |
Db      13525 ATACAAGTACTGATGATACTGTTCCTAATAATACTGTTCCAGTTACAGCTATTCCTTCTC 13584

Qy       1769 AATCATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATC- 1827
              || || |      |  ||||      |                 ||    ||||  | ||
Db      13585 TTGCAAATACTGGTGTTGACACTACTAGCAACAGTTTTTCCATTATGACCACTTCTTTCT 13644

Qy       1828 -----ATTAGAGAGCCACCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGT 1882
                  | ||   |  |   |||   ||||| | ||   |      ||   ||  |
Db      13645 CTGAAAGTACTAATGCTATGAACACTACTGTTATTATGGCAACTACTTCTCCTACAAGTA 13704

Qy       1883 TTGCTACTACCACAACTGTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGG 1942
              || |  | | |  ||    | || | |  | |      | |  |  | |    |  | |
Db      13705 CTGATGTTGCTAGCACAAATAATGATGCTTCTATGACAAATTTTCTTTTAGCTACAATGT 13764

Qy       1943 AACCACCAAACCCAACTGTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTA 2002
                |    ||  ||||   |  | || |          ||   || | ||  |  | |
Db      13765 CTGCTGGTAATATAACTAGTAATAGTATTTCCATAACAACTACTTCTTTTGGTAATAGTG 13824

Qy       2003 CTACCATTACCGCTCCACCTGGTGAAACTGATACCGTTCTTATCAGAGAGCCACCAAACC 2062
              | |  ||    || | |||  |  || |  |   | |  ||  | |         |
Db      13825 TTCCTTTTGTGACTACTCCTTCTCCAAGTACTGATGCTACTACTACAAGTAATAATACTA 13884

Qy       2063 ATACTGTCACTACTACTGAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTG 2122
              || ||| ||  |||||| | |||  | ||    |     | | |||   ||| |||||
Db      13885 ATCCTGGCATGACTACTTATTACCAGACTTCTCCTACCATTCCTACCCATACTCTTACTT 13944

Qy       2123 CACCACCTGGTGAAACCGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTA 2182
              |   ||| |    |   | |  || | | ||     ||| |||   ||||  ||||||
Db      13945 CTATTCCTAGCTCTATTACTTCTATTTTGAGCATGTTTCCAACAAGTAATACATTCACTA 14004

Qy       2183 CTACTGAA 2190
              ||  | ||
Db      14005 CTGATAAA 14012
```

```
RESULT 3
US-09-842-552-22/c
; Sequence 22, Application US/09842552
; Patent No. US20020055628A1
; GENERAL INFORMATION:
;  APPLICANT: The Regents of The University of California
;  TITLE OF INVENTION: MULTILOCUS REPETITIVE DNA SEQUENCES FOR GENOTYPING BACILLUS ANTHR
;  TITLE OF INVENTION:  RELATED BACTERIA
;  FILE REFERENCE: S-89,687
;  CURRENT APPLICATION NUMBER: US/09/842,552
;  CURRENT FILING DATE:  2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/199,911
;  PRIOR FILING DATE: 2000-04-26
;  NUMBER OF SEQ ID NOS: 106
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;  LENGTH: 2014
;  TYPE: DNA
;  ORGANISM: Bacillus anthracis
US-09-842-552-22

  Query Match             1.6%;  Score 62.2;  DB 10;  Length 2014;
  Best Local Similarity   42.3%;  Pred. No. 0.00092;
  Matches  402;  Conservative    0;  Mismatches  548;  Indels    1;  Gaps    1;

Qy       1354 ACTACAGTTACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAAC 1413
              || || | |  | || || |   ||||  |     |||  | |    |      |  | | |
Db       1356 ACAACTGGTTGTTCTTCTACGACTGGTGCTTCCTCTGCAACTGGTTGTTCTTCTACGATC 1297

Qy       1414 CATACTGTCACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACT 1473
              ||  |. | || | ||||  | ||  ||      ||  || | |  || |.|||  |  |
Db       1296 GATGTTTCCTCTGCAACTGGTTGTTCTTCTACGACCGATGTTTCCTCTGCAACTGGTTGT 1237

Qy       1474 GCTCCTCCAGGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACT 1533
              || || ||   ||| ||   ||  || |    |     | | |       ||    ||
Db       1236 TCTTCTGCAATCGGTGCTTCTTCTACAACTGATTGTCCTTCTACGACCGGTGCTTCTTCT 1177

Qy       1534 ACAACCGAGTATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGT 1593
              |||| ||    ||  ||| | |   |  | || |||| || | | || || || ||| |
Db       1176 ACAATCGGTGTTTCTTCTACAACTGATTGTTCTTCTACGACCGGTGCTTCTTCTGCAGCT 1117

Qy       1594 GGTACTGACTCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAATAT 1653
              |||  |   ||  |||          || |||       |    | || || |    |
Db       1116 GGTTGTTCTTCTACAATCGGTGTTTCTTCTACAACTGATTGTTCTTCTACGACCGGTGCT 1057

Qy       1654 TGGTCCCAATCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTGACTCA 1713
              |  ||   | ||       |  | || |  |    | || || ||   |    | |||
Db       1056 TCTTCTGCAACTGGTTGTTCTTCTACAATCGGTGTTTCTTCTACAACTGATTGTTCCTCC 997

Qy       1714 GTAATTATCAGAGAACCACCAAACCACAC-TGTCACTACTACTGAATACTGGTCACAATC 1772
              | |||||  | |   || |    || | |   | |  ||   |    | ||||
Db        996 GCAATTAAAACATCCGCAAAGCTTTGAACATCTTTTTTCGTTTGTTCATCAGCCACAGCA 937

Qy       1773 ATATGCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAG 1832
              | |||     | |    || |   |  |   |    |   ||      | || |||||
Db        936 ACATGTATGAATTCATTCTTTTGTTCAATTGCTTCATTTAACACCATTTCTTGTCATTTC 877

Qy       1833 AGAGCCACCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTAC 1892
                | |     |     || ||  ||  | |   |||  |     | |||||
Db        876 CTCAAGTGCTACAGGTTCCACTTCTTCTGGTGCTTCTGTTTCTGCAATTACTTCTACTTC 817

Qy       1893 CACAACTGTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAA 1952
              |       ||| | |  |||  |    |  | ||  ||     |     |
Db        816 TTCTGACTCTTCTGTTTCTGCAATTACTTCCACTTCTTCTAATTCTTCTAATTTTTCTGT 757

Qy       1953 CCCAACTGTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTAC 2012
              |  |  | ||  | |||   |     |||       | |||||| ||| | ||   | |
Db        756 TTCTGCAATTACTTCCACTTCTTCTAACTCTTCTGTTTCTGCAACTACTTCCACTTCTTC 697

Qy       2013 CGCTCCACCTGGTGAAACTGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCAC 2072
                | |  ||| |    | | |||      |       |  |  |      | |  | ||
Db        696 TAATTCTTCTGTTTCTGCAGTTACTTCCACTTCTTCTAACTCTTCTGTTTCTGCAATTAC 637

Qy       2073 TACTACTGAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGG 2132
              | |||||   |     |||      | ||| |  ||  | |||  | |||   |  |||
Db        636 TTCTACTTCTTCTGACTCTTCTGTTTCTGCAATTACTTCTACTTCTTCTGACTCTTCTGT 577

Qy       2133 TGAAACCGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATA 2192
              |    |  |||   |  |       ||  |  |       | |  | |||  |||||   |
Db        576 TTCCGCAATTACTTCTACTTCTTCTGACTCTTCTGTTTCTGCAATTACTTCTACTTCTTC 517

Qy       2193 CTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATAC 2252
                 |||       | ||| |  ||  | |||  | |||   |  | | |  | |   |||
Db        516 TGACTCTTCTGTTTCTGCAATTACTTCTACTTCTTCTGACTCTTCTGTTTCTGCAATTAC 457

Qy       2253 TGTTATCATTAGAGAGCCACCAAATCCAACAGTTACTACTACTGAATATTG 2303
              |  ||    |  ||  |  |  | |  || ||||| | |||   | | |
Db        456 TTCTACTTCTTCTGACTCTTCTGTTTCTGCAATTACTTCCACTTCTTCTGG 406
```

```
RESULT 4
US-09-864-761-2885
; Sequence 2885, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2885
;   LENGTH: 436
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AP000018.2
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.1
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
US-09-864-761-2885

  Query Match             1.5%;  Score 55.4;  DB 10;  Length 436;
  Best Local Similarity   48.6%;  Pred. No. 0.012;
  Matches  152;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy        965 CTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCTGTCA 1024
              |||  |  | |   |||    |||| | ||| ||    ||    |  | |||| |  |||
Db         68 CTATTGTCACTACTATTACCACTACCATTAGTACTACCACCACTATCATCACTACCATCA 127

Qy       1025 CTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAACCTA 1084
              |||||    |  ||| | | | |  |  ||  |  |  || |     |||       |||
Db        128 CTACTAATACTATCATTACCACTACTATTACCATTACCACGACTATCATTACTACTACTA 187

Qy       1085 TTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACTAAGA 1144
                |  |||| |  |  || |||     | |   |   ||||| | | | | | ||  |
Db        188 CCACCTCCACCATTACTACTACTAATGCTATCATTACTACTACTGCTTCTATAATTACTA 247

Qy       1145 CTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACTACCACAA 1204
              || |  |   |      || |||||  || | |     |  |   |  ||||| ||  |
Db        248 CTCCTACTCCTACCATTACTACTACTAGTAGTATCACCATTACTACTACTACTATCATTA 307

Qy       1205 CTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAATCCAACTG 1264
              || ||||||        ||    |  || ||||| || || ||    |  | | | ||||
Db        308 CTATTACCACCACCATTACTACTATCATTACTACTACTACCACCACCATTACTACTACTG 367

Qy       1265 ATTCAATTGACAC 1277
                |  |||  ||
Db        368 CTGTCATTACTAC 380


RESULT 5
US-09-216-393-340
; Sequence 340, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393
;  CURRENT FILING DATE: 1998-12-18
;  EARLIER APPLICATION NUMBER: 08/994,825
;  EARLIER FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 364
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 340
;   LENGTH: 867
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(867)
US-09-216-393-340

  Query Match             1.4%;  Score 54.6;  DB 10;  Length 867;
  Best Local Similarity   45.5%;  Pred. No. 0.028;
  Matches  195;  Conservative    0;  Mismatches  234;  Indels    0;  Gaps    0;

Qy       3017 CTGATTTGACTAGTGAACCAACAAATACCAGAGAACAACCAACTACATTATCAACTACTT 3076
              |||  || | |||   |   |     | ||||   ||  ||||||    |||| | ||
Db        395 CTGCGTTCATTAGGCCAAGGAGGTCAAAGAGAGGGAAAAAGACTACAACTACAACCAGTT 454

Qy       3077 CAAACTCCATCACTGAAGATATCACCACATCTCAACCTACAGGTGATAATGGAGACAATA 3136
              ||  | | |  ||       |  |||||  |   | ||||   |  ||       ||  |
Db        455 CATCCACAAGTACGAGTACAACGACCACGACATCAACTACCACTACTACCACTACCACCA 514

Qy       3137 CTTCATCAACCAATCCAGTTCCAACTGTGGCAACAAGTACTTTAGCATCTGCAAGTGAAG 3196
              || |  | || | | ||  | | ||     ||||| |||   | |  |  ||| |  |
Db        515 CTACGACTACTACTACAACTACGACACCAACAACAACTACAACAACCACAACAACTACAC 574

Qy       3197 AAGACAACAAAAGCGGTTCTCATGAATCAGCATCCACAAGTTTGAAACCAAGTATGGGTG 3256
                |  | |  |       |       | |  || | ||||    || | |||  |
Db        575 CAACAACAACGACAACAACCACAACAACTACACCAACAACAACGACAACAACCACAACAA 634

Qy       3257 AAAATTCTGGATTAACTACTTCTACTGAAATTGAAGCTACAACAACCAGTCCTACAGAAG 3316
                |   |  | ||| ||  | ||   ||    | |||||||  ||   | | |||
```

```
Db        635 CTACACCAACAACAACGACAACAACCACAACGCCAACTACAACGACATCTACGACAACCA 694

Qy       3317 CTCCATCACCTGCTGTTTCTTCTGGTACTGATGTAACTACTGAACCAACTGATACTAGAG 3376
              || |  |  |  |   | || ||    |   |  ||| ||   | | || ||  | | |
Db        695 CTACGACTACCACAACTACTACTACACCAACTACAACAACGACAACCACGGAACCAACAA 754

Qy       3377 AACAACCTACTACATTATCAACTACTTCAAAAACAAACAGTGAACTGGTTGCTACTACAC 3436
                | | || ||      | | ||  |||  |||| ||       |  | | || ||
Db        755 CTACAACAACAACCACGGAACCAACCACAACTACAAGCACAACGACGACTACGACAACTA 814

Qy       3437 AAGCTACTA 3445
                | | ||||
Db        815 CAACGACTA 823


RESULT 6
US-09-216-393-342
; Sequence 342, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393
;  CURRENT FILING DATE: 1998-12-18
;  EARLIER APPLICATION NUMBER: 08/994,825
;  EARLIER FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 364
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 342
;   LENGTH: 867
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
US-09-216-393-342

  Query Match             1.4%;  Score 54.6;  DB 10;  Length 867;
  Best Local Similarity   45.5%;  Pred. No. 0.028;
  Matches  195;  Conservative    0;  Mismatches  234;  Indels    0;  Gaps    0;

Qy       3017 CTGATTTGACTAGTGAACCAACAAATACCAGAGAACAACCAACTACATTATCAACTACTT 3076
              |||  || | |||   |   |      |  ||||   ||   ||||||    |||| | ||
Db        395 CTGCGTTCATTAGGCCAAGGAGGTCAAAGAGAGGGAAAAAGACTACAACTACAACCAGTT 454

Qy       3077 CAAACTCCATCACTGAAGATATCACCACATCTCAACCTACAGGTGATAATGGAGACAATA 3136
              ||  | | |  ||       |  |||||  |   | ||||   |  ||       ||  |
Db        455 CATCCACAAGTACGAGTACAACGACCACGACATCAACTACCACTACTACCACTACCACCA 514

Qy       3137 CTTCATCAACCAATCCAGTTCCAACTGTGGCAACAAGTACTTTAGCATCTGCAAGTGAAG 3196
              || |  | || | | ||  | | ||     ||||| |||   | |  |  ||| |  |
Db        515 CTACGACTACTACTACAACTACGACACCAACAACAACTACAACAACCACAACAACTACAC 574

Qy       3197 AAGACAACAAAAGCGGTTCTCATGAATCAGCATCCACAAGTTTGAAACCAAGTATGGGTG 3256
                |   | | |       |       | |  || | ||||    || | |||  |
Db        575 CAACAACAACGACAACAACCACAACAACTACACCAACAACAACGACAACAACCACAACAA 634

Qy       3257 AAAATTCTGGATTAACTACTTCTACTGAAATTGAAGCTACAACAACCAGTCCTACAGAAG 3316
               |   |   | ||| ||  | ||   ||    | |||||| ||   | | |||
Db        635 CTACACCAACAACAACGACAACAACCACAACGCCAACTACAACGACATCTACGACAACCA 694

Qy       3317 CTCCATCACCTGCTGTTTCTTCTGGTACTGATGTAACTACTGAACCAACTGATACTAGAG 3376
              || |  |  |  |   | || ||    |   |  ||| ||   | | || ||  | | |
Db        695 CTACGACTACCACAACTACTACTACACCAACTACAACAACGACAACCACGGAACCAACAA 754

Qy       3377 AACAACCTACTACATTATCAACTACTTCAAAAACAAACAGTGAACTGGTTGCTACTACAC 3436
                | | || ||      | | ||  |||  |||| ||       |  | | || ||
Db        755 CTACAACAACAACCACGGAACCAACCACAACTACAAGCACAACGACGACTACGACAACTA 814

Qy       3437 AAGCTACTA 3445
                | | ||||
Db        815 CAACGACTA 823


RESULT 7
US-09-216-393-343
; Sequence 343, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393
;  CURRENT FILING DATE: 1998-12-18
;  EARLIER APPLICATION NUMBER: 08/994,825
;  EARLIER FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 364
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 343
;   LENGTH: 1397
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (238)..(1104)
US-09-216-393-343

  Query Match             1.4%;  Score 54.6;  DB 10;  Length 1397;
  Best Local Similarity   45.5%;  Pred. No. 0.038;
  Matches  195;  Conservative    0;  Mismatches  234;  Indels    0;  Gaps    0;

Qy       3017 CTGATTTGACTAGTGAACCAACAAATACCAGAGAACAACCAACTACATTATCAACTACTT 3076
              |||  || | |||   |   |      |  ||||   ||   ||||||    |||| | ||
Db        632 CTGCGTTCATTAGGCCAAGGAGGTCAAAGAGAGGGAAAAAGACTACAACTACAACCAGTT 691

Qy       3077 CAAACTCCATCACTGAAGATATCACCACATCTCAACCTACAGGTGATAATGGAGACAATA 3136
              ||  | | |  ||       |  |||||  |   | ||||   |  ||       ||  |
Db        692 CATCCACAAGTACGAGTACAACGACCACGACATCAACTACCACTACTACCACTACCACCA 751

Qy       3137 CTTCATCAACCAATCCAGTTCCAACTGTGGCAACAAGTACTTTAGCATCTGCAAGTGAAG 3196
              || |  | || | | ||  | | ||     ||||| |||   | |  |  ||| |  |
Db        752 CTACGACTACTACTACAACTACGACACCAACAACAACTACAACAACCACAACAACTACAC 811

Qy       3197 AAGACAACAAAAGCGGTTCTCATGAATCAGCATCCACAAGTTTGAAACCAAGTATGGGTG 3256
                |   | | |       |       | |  || | ||||    || | |||  |
Db        812 CAACAACAACGACAACAACCACAACAACTACACCAACAACAACGACAACAACCACAACAA 871

Qy       3257 AAAATTCTGGATTAACTACTTCTACTGAAATTGAAGCTACAACAACCAGTCCTACAGAAG 3316
               |   |   | ||| ||  | ||   ||    | |||||| ||   | | |||
Db        872 CTACACCAACAACAACGACAACAACCACAACGCCAACTACAACGACATCTACGACAACCA 931

Qy       3317 CTCCATCACCTGCTGTTTCTTCTGGTACTGATGTAACTACTGAACCAACTGATACTAGAG 3376
              || |  |  |  |   | || ||    |   |  ||| ||   | | || ||  | | |
Db        932 CTACGACTACCACAACTACTACTACACCAACTACAACAACGACAACCACGGAACCAACAA 991

Qy       3377 AACAACCTACTACATTATCAACTACTTCAAAAACAAACAGTGAACTGGTTGCTACTACAC 3436
                | | || ||      | | ||  |||  |||| ||       |  | | || ||
Db        992 CTACAACAACAACCACGGAACCAACCACAACTACAAGCACAACGACGACTACGACAACTA 1051

Qy       3437 AAGCTACTA 3445
                | | ||||
Db       1052 CAACGACTA 1060


RESULT 8
US-09-216-393-345/c
; Sequence 345, Application US/09216393
; Patent No. US20010014447A1
; GENERAL INFORMATION:
;  APPLICANT: Milhausen, Michael James
;  TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND
;  TITLE OF INVENTION: USES THEREOF
;  FILE REFERENCE: TX-1-C2
;  CURRENT APPLICATION NUMBER: US/09/216,393
;  CURRENT FILING DATE: 1998-12-18
;  EARLIER APPLICATION NUMBER: 08/994,825
```

```
;  EARLIER FILING DATE: 1997-12-19
;  NUMBER OF SEQ ID NOS: 364
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 345
;   LENGTH: 1397
;   TYPE: DNA
;   ORGANISM: Toxoplasma gondii
US-09-216-393-345

  Query Match             1.4%;  Score 54.6;  DB 10;  Length 1397;
  Best Local Similarity   45.5%;  Pred. No. 0.038;
  Matches  195;  Conservative    0;  Mismatches  234;  Indels    0;  Gaps    0;

Qy       3017 CTGATTTGACTAGTGAACCAACAAATACCAGAGAACAACCAACTACATTATCAACTACTT 3076
              |||  || | |||   |   |     |  ||||   ||   ||||||    |||| | ||
Db        766 CTGCGTTCATTAGGCCAAGGAGGTCAAAGAGAGGGAAAAAGACTACAACTACAACCAGTT 707

Qy       3077 CAAACTCCATCACTGAAGATATCACCACATCTCAACCTACAGGTGATAATGGAGACAATA 3136
              ||  | | |  ||       |  |||||  |   | ||||   |  ||       ||  |
Db        706 CATCCACAAGTACGAGTACAACGACCACGACATCAACTACCACTACTACCACTACCACCA 647

Qy       3137 CTTCATCAACCAATCCAGTTCCAACTGTGGCAACAAGTACTTTAGCATCTGCAAGTGAAG 3196
              || |  | || | | ||  | | ||     ||||||| |||   | |  |  ||| |  |
Db        646 CTACGACTACTACTACAACTACGACACCAACAACAACTACAACAACCACAACAACTACAC 587

Qy       3197 AAGACAACAAAAGCGGTTCTCATGAATCAGCATCCACAAGTTTGAAACCAAGTATGGGTG 3256
              |   |  |  |      |      | |  || | ||||    || | |||  |
Db        586 CAACAACAACGACAACAACCACAACAACTACACCAACAACAACGACAACAACCACAACAA 527

Qy       3257 AAAATTCTGGATTAACTACTTCTACTGAAATTGAAGCTACAACAACCAGTCCTACAGAAG 3316
              |   |   | ||| || ||  ||    | |||||| ||   | | |||
Db        526 CTACACCAACAACAACGACAACAACCACAACGCCAACTACAACGACATCTACGACAACCA 467

Qy       3317 CTCCATCACCTGCTGTTTCTTCTGGTACTGATGTAACTACTGAACCAACTGATACTAGAG 3376
              || |  |  |  |   | || ||    |   |  ||| ||   | | || ||  | | |
Db        466 CTACGACTACCACAACTACTACTACACCAACTACAACAACGACAACCACGGAACCAACAA 407

Qy       3377 AACAACCTACTACATTATCAACTACTTCAAAAACAAACAGTGAACTGGTTGCTACTACAC 3436
                | | || ||      | | ||  |||  |||| ||       |  | | || ||
Db        406 CTACAACAACAACCACGGAACCAACCACAACTACAAGCACAACGACGACTACGACAACTA 347

Qy       3437 AAGCTACTA 3445
              | | ||||
Db        346 CAACGACTA 338


RESULT 9
US-09-801-368-107
; Sequence 107, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
;  APPLICANT: Busby, Robert
;  APPLICANT:  Cali, Brian
;  APPLICANT:  Hecht, Peter
;  APPLICANT:  Holtzman, Doug
;  APPLICANT:  Madden, Kevin
;  APPLICANT:  Maxon, Mary
;  APPLICANT:  Milne, Todd
;  APPLICANT:  No. US20020128250A1man, Thea
;  APPLICANT:  Royer, John
;  APPLICANT:  Salama, Sofie
;  APPLICANT:  Sherman, Amir
;  APPLICANT:  Silva, Jeff
;  APPLICANT:  Summers, Eric
;  TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
;  FILE REFERENCE: 109272.147
;  CURRENT APPLICATION NUMBER: US/09/801,368
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: US 09/487,558
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: US 60/160,587
;  PRIOR FILING DATE: 1999-10-20
;  NUMBER OF SEQ ID NOS: 440

;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 107
;   LENGTH: 4104
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-107

  Query Match             1.4%;  Score 52.8;  DB 10;  Length 4104;
  Best Local Similarity   43.7%;  Pred. No. 0.19;
  Matches  335;  Conservative    0;  Mismatches  422;  Indels    9;  Gaps    2;

Qy       2948 CTTCCACTGAAATTGAAATTGTAACAACCAGTTCTACTAAAGTTTTACCACCTGTCGTTT 3007
              ||||  ||  | |   || |  | | | |  | |||||  ||   |   |  ||   ||
Db       1901 CTTCTGCTCCAGTACCAACTCCATCCAGCTCTACTACTGAAAGCTCTTCTGCTCCAGTAC 1960

Qy       3008 CTTCTAATACTGATTTGACTAGTGAACCAACAAATACCAGAGAACAACCAACTACATTAT 3067
              |  |      |     |  |||| ||||     |   | |   || || | |  |  |    |
Db       1961 CAACCCCATCCAGCTCTACTACTGAAAGCTCTTCTGCTCCAGTACCAACTCCATCCAGCT 2020

Qy       3068 CAACTACTTCAAACTCCATCACTGAAGATATCACCACATCTCAACCTACAGGTGATAATG 3127
              | ||||||  || |||     ||  ||  | || | |  |       |   |  |  |
Db       2021 CTACTACTGAAAGCTCTTCTGCTCCAGTAACCAGCTCTACCACTGAAAGCTCTTCTGCTC 2080

Qy       3128 GAGACAATACTTCATCAACCAATCCAGTTCCAACTGTGGCAACAAGTACTTTAGCATCTG 3187
              ||  |  |  ||  | ||  |      | |  ||   | | ||| | | |     |||
Db       2081 CAGTAACCAGCTCTACCACTGAAAGCTCTTCTGCTCCAGTACCAACTCCATCCAGCTCTA 2140

Qy       3188 CAAGTGAAGAAGACAACAAAAGCGGTTCTCATGAATCAGCATCCACAAGTTTGAAACCAA 3247
              | | ||||              |   |  | |||    || || | |   |   |
Db       2141 CTACTGAAAGCTCTTCTGCTCCAGTACCAACTCCATCCAGCTCTACTACTGAAAGCTCTT 2200

Qy       3248 GTATGGGTGAAAATTCTGGATTAACTACTTCTACTGAAA------TTGAAGCTACAACAA 3301
              |      | |    ||  ||  |  || |||||||||       ||   |  ||| |
Db       2201 CTGCTCCAGTACCAACTCCATCCAGCTCTACTACTGAAAGCTCTTCTGCTCCAGTAACCA 2260

Qy       3302 CCAGTCCTACAGAAGCTCCATCACCTGCTGTTTCTTCTGGTACTGATGTAACTACTGAAC 3361
              |  | | ||  |||    | ||  || | ||  |  ||     |        |||||||||
Db       2261 GCTCTACCACTGAAAGCTCTTCTGCTCCAGTACCAACTCCATCCAGCTCTACTACTGAAA 2320

Qy       3362 CAACTGATACTAGAGAACAACCTACTACATTATCAACTACTTCAAAAACAAACAGTGAAC 3421
              ·||  | ||  ||  | | || |  |   || |||||  ||   |       |  |
Db       2321 GCTCTTCTGCTCCAGTGCCAACTCCATCCAGCTCTACTACCGAAAGCTCTTCTGCTCCAG 2380

Qy       3422 TGGTTGCTACTACACAAGCTACTAATGAAAATGGTGGTAAATCTCCATCAACTGATTTAA 3481
              |      |  |  ||    |||||| |||||     ||    | |   | ||||    |
Db       2381 TACCAACCCCATCAAGCTCTACTACTGAAAGCTCTGTAGCACCAGTACCAACCCCATCTT 2440

Qy       3482 CATCAAGCTTGACAACAGGCACCTCAGCATCTACAAGTGCTAATAGCGAACTTGTTACTA 3541
              | || |||   |  ||   | || |  ||||| |||   |    |||        |  |
Db       2441 CCTCTAGCAACATCACTTCCTCCGCTCCATCTTCAACCCCATTCAGCTCTAGCACTGAAA 2500

Qy       3542 GTGGATCTGTTACTGGTGGAGCTGTTGCCAGTGCTTCAAATGATCAATCACATTCTAC-- 3599
              |    |||||| | |    | |     | ||  || | | |||    ||   | || |
Db       2501 GCTCTTCTGTTCCAGTACCAACCCCATCAAGCTCTACTACTGAAAGCTCTTCTGCTCCAG 2560

Qy       3600 -TTCTGTTACCAACAGCAACAGCATTGTATCTAATACCCCACAAACTACATTGAGTCAAC 3658
               ||     ||| ||   | |||  |||| |       |||    |  | |    |
Db       2561 TATCCAGCTCCACCACTGAAAGCTCTGTAGCACCAGTACCAACCCCATCTTCCTCTAGCA 2620

Qy       3659 AAGTTACCTCATCCTCACCTTCAACCAACACATTCATTGCTTCTAC 3704
              |  | || || ||| | || ||  | |   ||||||   || ||||
Db       2621 ACATCACTTCCTCCGCTCCATCTTCAATTCCATTCAGCTCTACTAC 2666


RESULT 10
US-10-073-256-54/c
; Sequence 54, Application US/10073256
; Patent No. US20020120408A1
; GENERAL INFORMATION:
;  APPLICANT: Kreiswirth, Barry N
;  APPLICANT:  Nadich, Steven M
;  TITLE OF INVENTION: System and Method for Tracking and Controlling Infections
;  FILE REFERENCE: 19124.0002
```

```
;  CURRENT APPLICATION NUMBER: US/10/073,256
;  CURRENT FILING DATE:  2002-02-13
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54
;   LENGTH: 1059
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-10-073-256-54

  Query Match             1.4%;  Score 52.2;  DB 12;  Length 1059;
  Best Local Similarity   41.6%;  Pred. No. 0.11;
  Matches  333;  Conservative    0;  Mismatches  468;  Indels    0;  Gaps    0;

Qy       1303 AATCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTT 1362
Db       1010 ACTTCACCTTTAGGATTTGATGGTGCTTTCTGTTCATTATTTGGTGGTGTAACTCTTGAA 951

Qy       1363 ACTGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTC 1422
Db        950 TCGGAGTCACTATCTGAGTCTGAGTCGCTATCTGAATCCGAGTCGCTATCCGAGTCTGAG 891

Qy       1423 ACTACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCA 1482
Db        890 TCGCTATCTGAGTCTGAATCGCTGTCTGAGTCTGAGTCGCTATCCGAGTCTGAGTCGCTG 831

Qy       1483 GGTGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAG 1542
Db        830 TCTGAATCTGAATCACTGTCTGAATCCGAATCGCTATCTGAATCTGAATCGCTATCCGAG 771

Qy       1543 TATTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGAC 1602
Db        770 TCTGAGTCGCTGTCTGAATCTGAATCGCTGTCTGAGTCCGAATCGCTATCTGAATCTGAG 711

Qy       1603 TCAGTAATTATCAGAGAACCTCCAAACCCAACTGTCACCACCACTGAATATTGGTCCCAA 1662
Db        710 TCGCTGTCTGAGTCTGAATCGCTATCTGAATCTGAGTCGCTATCTGAGTCTGAGTCGCTG 651

Qy       1663 TCTTACGCAACCACAACTACTGTGACTGCTCCTCCAGGAGGCACTGACTCAGTAATTATC 1722
Db        650 TCTGAGTCTGAGTCGCTGTCTGAGTCTGAATCGCTATCTGAATCTGAGTCGCTGTCTGAG 591

Qy       1723 AGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATATGCCACC 1782
Db        590 TCTGAGTCGCTATCTGAGTCTGAGTCGCTGTCTGAATCTGAGTCGCTGTCTGAATCTGAA 531

Qy       1783 ACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAGCCACCA 1842
Db        530 TCGCTATCTGAGTCTGAATCGCTATCTGAGTCTGAATCACTGTCTGAGTCCGAGTCACTG 471

Qy       1843 AACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACAACTGTA 1902
Db        470 TCTGAATCTGACTCACTATCTGATTCTGAGTCGCTATCTGATTCTGAGTCGCTGTCTGAA 411

Qy       1903 ACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCAACTGTC 1962
Db        410 TCTGAATCACTGTCTGAATCCGAATCGCTATCTGAATCTGAATCGCTATCTGAATCTGAA 351

Qy       1963 ACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCTCCACCT 2022
Db        350 TCGCTATCTGAATCTGAGTCGCTATCTGATTCTGAATCGCTATCTGATTCTGAGTCGCTG 291

Qy       2023 GGTGAAACTGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAA 2082
Db        290 TCTGAATCTGAATCACTGTCTGAATCCGAATCGCTATCTGATTCTGAGTCGCTATCTGAA 231

Qy       2083 TACTGGTCTCAATCATATGCT 2103
Db        230 CCTGAGTCGCTGTCTGAGCCT 210


RESULT 11
US-09-842-552-79/c
; Sequence 79, Application US/09842552
; Patent No. US20020055628A1
; GENERAL INFORMATION:
;  APPLICANT: The Regents of The University of California
;  TITLE OF INVENTION: MULTILOCUS REPETITIVE DNA SEQUENCES FOR GENOTYPING BACILLUS AN
;  TITLE OF INVENTION:  RELATED BACTERIA
;  FILE REFERENCE: S-89,687
;  CURRENT APPLICATION NUMBER: US/09/842,552
;  CURRENT FILING DATE:  2001-04-23
;  PRIOR APPLICATION NUMBER: US 60/199,911
;  PRIOR FILING DATE: 2000-04-26
;  NUMBER OF SEQ ID NOS: 106
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 79
;   LENGTH: 2015
;   TYPE: DNA
;   ORGANISM: Bacillus anthracis
US-09-842-552-79

  Query Match             1.4%;  Score 52;  DB 10;  Length 2015;
  Best Local Similarity   48.3%;  Pred. No. 0.18;
  Matches  145;  Conservative    0;  Mismatches  155;  Indels    0;  Gaps    0;

Qy       1305 TCCAACTGTTAGTACTACTGAATATTGGTCTCAGTCCTTTGCTACAACCACTACAGTTAC 1364
Db       1370 TTCTTCTTTTTGTACAACTGGTTGTTCTTCTACGACTGGTGCTTCCTCTGCAACTGGTTG 1311

Qy       1365 TGCTCCTCCAGGTGGTACCGATACTGTGATTATCAGAGAGCCACCAAACCATACTGTCAC 1424
Db       1310 TTCTTCTACGATCGATGTTTCCTCTGCAACTGGTTGTTCTTCTACGACCGATGTTTCCTC 1251

Qy       1425 TACTACTGAATATTGGTCACAATCCTTTGCTACTACTACTACTGTTACTGCTCCTCCAGG 1484
Db       1250 TGCAACTGGTTGTTCTTCTGCAATCGGTGCTTCTTCTACAACTGATTGTCCTTCTACGAC 1191

Qy       1485 TGGTACTGACTCAGTAATTATCAGAGAACCACCAAATCCAACTGTCACTACAACCGAGTA 1544
Db       1190 CGGTGCTTCTTCTACAATCGGTGTTTCTTCTACAACTGATTGTTCTTCTACGACCGGTGC 1131

Qy       1545 TTGGTCTCAATCCTTTGCTACTACTACTACAGTTACTGCTCCTCCAGGTGGTACTGACTC 1604
Db       1130 TTCTTCTGCAGCTGGTTGTTCTTCTACAATCGGTGTTTCTTCTACAACTGATTGTTCTTC 1071


RESULT 12
US-10-025-380-1058
; Sequence 1058, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole L.
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  APPLICANT:  Skeiky, Yasir A. W.
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick Thomas S.
;  APPLICANT:  Carter, Darrick
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C14
;  CURRENT APPLICATION NUMBER: US/10/025,380
;  CURRENT FILING DATE:  2001-12-19
;  NUMBER OF SEQ ID NOS: 1129
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1058
```

```
;   LENGTH: 15720
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-025-380-1058

  Query Match             1.3%;  Score 50.8;  DB 9;  Length 15720;
  Best Local Similarity   43.8%;  Pred. No. 1.2;
  Matches  269;  Conservative    0;  Mismatches  342;  Indels    3;  Gaps    1;

Qy       1717 ATTATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATAT 1776
              | |  |||    ||| ||| |  |||| || | || |   |  |    ||    |  |
Db       4756 ACTCCCAGCCCTCCAACAACCACCACTCCCAGTCCTCCAACAATCACCACAACCACCCCT 4815

Qy       1777 GCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAG 1836
              |  | || ||||||   |   | ||| ||      ||   |||   |       |
Db       4816 CCACCAACCACCACTCCCAGCCCTCCAACAACGACCACAACCACCCCTCCACCAACCACC 4875

Qy       1837 CCACCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACA 1896
              | || | ||   ||   || |||||           | | | |     |||| |||
Db       4876 ACTCCCAGCC---CTCCAACGACTACACCCATCACTCCACCAACCAGCACTACCACCCTT 4932

Qy       1897 ACTGTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCA 1956
              |   |||     | || ||     |    ||    |  |  |     |||||||  | |
Db       4933 CCACCAACCACCACTCCCAGCCCTCCACCAACCACCACAACCACCCCTCCACCAACCACC 4992

Qy       1957 ACTGTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCT 2016
              |||  || | || |   |  |    |||       | |||  || || || |  ||| ||
Db       4993 ACTCCCAGCCCTCCAACAACCACCACTCCCAGCCCTCCAATAACCACCACAACCACCCCT 5052

Qy       2017 CCACCTGGTGAAACTGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACT 2076
              |||||       |||  | |  ||  || |      | || | || | |   || ||
Db       5053 CCACCAACCACCACTCCCAGCTCTCCAATAACCACCACTCCCAGCCCTCCAACAACCACC 5112

Qy       2077 ACTGAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAA 2136
              |      |   || | | |     | | | | |  |  | ||  |  || | |
Db       5113 ATGACCACCCCTTCACCAACCACCACCCCCAGCTCTCCAATAACCACCACAACCACCCCT 5172

Qy       2137 ACCGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGG 2196
              ||   ||      |  |||    ||||||| |   |    |||  || |   |  |
Db       5173 TCCTCAACTACCACTCCCAGCCCTCCACCAACCACCATGACCACCCCTTCACCAACCACC 5232

Qy       2197 TCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTT 2256
              |||       | | ||||||||||     ||    ||||||      ||    |    |
Db       5233 ACTCCCAGCCCTCCAACAACCACCATGACCACCCTTCCACCAACCACCACTTCCAGCCCT 5292

Qy       2257 ATCATTAGAGAGCCACCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTT 2316
              | |  |     || | |  |||| || | ||| || ||  || ||   ||  |||
Db       5293 CTAACAACTACTCCTCTACCTCCATCAATAACTCCTCCTACATTTTCACCATTCTCAACG 5352

Qy       2317 GCCACAACCACCAC 2330
              | ||  | || ||
Db       5353 ACAACCCCTACTAC 5366


RESULT 13
US-09-922-217-1058
; Sequence 1058, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole Lynn
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C13
;  CURRENT APPLICATION NUMBER: US/09/922,217
;  CURRENT FILING DATE:  2001-08-03
;  NUMBER OF SEQ ID NOS: 1124
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1058
;   LENGTH: 15720
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-922-217-1058

  Query Match             1.3%;  Score 50.8;  DB 10;  Length 15720;
  Best Local Similarity   43.8%;  Pred. No. 1.2;
  Matches  269;  Conservative    0;  Mismatches  342;  Indels    3;  Gaps    1;

Qy       1717 ATTATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATAT 1776
              | |  |||    ||| ||| |  |||| || | || |   |  |    ||    |  |
Db       4756 ACTCCCAGCCCTCCAACAACCACCACTCCCAGTCCTCCAACAATCACCACAACCACCCCT 4815

Qy       1777 GCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAG 1836
              |  | || ||||||   |   | ||| ||      ||   |||   |       |
Db       4816 CCACCAACCACCACTCCCAGCCCTCCAACAACGACCACAACCACCCCTCCACCAACCACC 4875

Qy       1837 CCACCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACA 1896
              | || | ||   ||   || |||||           | | | |     |||| |||
Db       4876 ACTCCCAGCC---CTCCAACGACTACACCCATCACTCCACCAACCAGCACTACCACCCTT 4932

Qy       1897 ACTGTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCA 1956
              |   |||     | || ||     |    ||    |  |  |     |||||||  | |
Db       4933 CCACCAACCACCACTCCCAGCCCTCCACCAACCACCACAACCACCCCTCCACCAACCACC 4992

Qy       1957 ACTGTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCT 2016
              |||  || | || |   |  |    |||       | |||  || || || |  ||| ||
Db       4993 ACTCCCAGCCCTCCAACAACCACCACTCCCAGCCCTCCAATAACCACCACAACCACCCCT 5052

Qy       2017 CCACCTGGTGAAACTGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACT 2076
              |||||       |||  | |  ||  || |      | || | || | |   || ||
Db       5053 CCACCAACCACCACTCCCAGCTCTCCAATAACCACCACTCCCAGCCCTCCAACAACCACC 5112

Qy       2077 ACTGAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAA 2136
              |      |   || | | |     | | | | |  |  | ||  |  || | |
Db       5113 ATGACCACCCCTTCACCAACCACCACCCCCAGCTCTCCAATAACCACCACAACCACCCCT 5172

Qy       2137 ACCGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGG 2196
              ||   ||      |  |||    ||||||| |   |    |||  || |   |  |
Db       5173 TCCTCAACTACCACTCCCAGCCCTCCACCAACCACCATGACCACCCCTTCACCAACCACC 5232

Qy       2197 TCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTT 2256
              |||       | | ||||||||||     ||    ||||||      ||    |    |
Db       5233 ACTCCCAGCCCTCCAACAACCACCATGACCACCCTTCCACCAACCACCACTTCCAGCCCT 5292

Qy       2257 ATCATTAGAGAGCCACCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTT 2316
              | |  |     || | |  |||| || | ||| || ||  || ||   ||  |||
Db       5293 CTAACAACTACTCCTCTACCTCCATCAATAACTCCTCCTACATTTTCACCATTCTCAACG 5352

Qy       2317 GCCACAACCACCAC 2330
              | ||  | || ||
Db       5353 ACAACCCCTACTAC 5366


RESULT 14
US-09-833-263-1058
; Sequence 1058, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Meagher, Madeleine J.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C12
;  CURRENT APPLICATION NUMBER: US/09/833,263
```

```
;  CURRENT FILING DATE:  2001-04-10
;  NUMBER OF SEQ ID NOS: 1093
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1058
;   LENGTH: 15720
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-833-263-1058

  Query Match             1.3%;  Score 50.8;  DB 10;  Length 15720;
  Best Local Similarity   43.8%;  Pred. No. 1.2;
  Matches  269;  Conservative    0;  Mismatches  342;  Indels    3;  Gaps    1;

Qy       1717 ATTATCAGAGAACCACCAAACCACACTGTCACTACTACTGAATACTGGTCACAATCATAT 1776
              | |  |||    ||| ||| |  |||| || | || |   |  |    ||    |   |
Db       4756 ACTCCCAGCCCTCCAACAACCACCACTCCCAGTCCTCCAACAATCACCACAACCACCCCT 4815

Qy       1777 GCCACCACTACCACTGTAACTGCACCACCAGGTGGTACTGACACTGTTATCATTAGAGAG 1836
              |  | || |||||   |   | ||| ||      ||   |||  |       |
Db       4816 CCACCAACCACCACTCCCAGCCCTCCAACAACGACCACAACCACCCCTCCACCAACCACC 4875

Qy       1837 CCACCAAACCACACTGTCACTACTACTGAGTATTGGTCTCAATCGTTTGCTACTACCACA 1896
              | || | ||   ||   || |||||           | | | |     |||| |||
Db       4876 ACTCCCAGCC-:-CTCCAACGACTACACCCATCACTCCACCAACCAGCACTACCACCCTT 4932

Qy       1897 ACTGTAACTGGTCCACCAAGTGGCACTGATACTGTTATCATTAGGGAACCACCAAACCCA 1956
              |   |||     | || ||     |    ||    |  |  |     |||||| | |
Db       4933 CCACCAACCACCACTCCCAGCCCTCCACCAACCACCACAACCACCCCTCCACCAACCACC 4992

Qy       1957 ACTGTCACCACTACTGAATACTGGTCTCAATCATATGCAACCACTACTACCATTACCGCT 2016
              |||  || | || |   |  |    |||       | |||  || || || |  ||| ||
Db       4993 ACTCCCAGCCCTCCAACAACCACCACTCCCAGCCCTCCAATAACCACCACAACCACCCCT 5052

Qy       2017 CCACCTGGTGAAACTGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACT 2076
              |||||       |||   | |  ||  || |      | || | || | |    || ||
Db       5053 CCACCAACCACCACTCCCAGCTCTCCAATAACCACCACTCCCAGCCCTCCAACAACCACC 5112

Qy       2077 ACTGAATACTGGTCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCTGGTGAA 2136
              |       |   || | | |     |  | | | | |  | ||  |  || |
Db       5113 ATGACCACCCCTTCACCAACCACCACCCCCAGCTCTCCAATAACCACCACAACCACCCCT 5172

Qy       2137 ACCGATACCGTTCTTATCAGAGAGCCACCAAACCATACTGTCACTACTACTGAATACTGG 2196
              ||   ||      |  |||    |||||||  |   |    |||  || |   |  |
Db       5173 TCCTCAACTACCACTCCCAGCCCTCCACCAACCACCATGACCACCCCTTCACCAACCACC 5232

Qy       2197 TCTCAATCATATGCTACAACCACCACTGTTACTGCACCACCAGGTGGTACCGATACTGTT 2256
              |||       | | ||||||||||     ||    ||||||      ||    |    |
Db       5233 ACTCCCAGCCCTCCAACAACCACCATGACCACCCTTCCACCAACCACCACTTCCAGCCCT 5292

Qy       2257 ATCATTAGAGAGCCACCAAATCCAACAGTTACTACTACTGAATATTGGTCACAATCATTT 2316
              | |  |     || | |  |||| || | ||| || ||  || ||   ||   |||
Db       5293 CTAACAACTACTCCTCTACCTCCATCAATAACTCCTCCTACATTTTCACCATTCTCAACG 5352

Qy       2317 GCCACAACCACCAC 2330
              | ||  | || ||
Db       5353 ACAACCCCTACTAC 5366


RESULT 15
US-10-077-584-3
; Sequence 3, Application US/10077584
; Publication No. US20030073610A1
; GENERAL INFORMATION:
;  APPLICANT: LINDQUIST, SUSAN
;  APPLICANT:  KROBITSCH, SYLVIA
;  APPLICANT:  OUTEIRO, TIAGO F.
;  TITLE OF INVENTION: YEAST SCREENS FOR THE TREATMENT OF HUMAN DISEASE
;  FILE REFERENCE: ARCD:367US
;  CURRENT APPLICATION NUMBER: US/10/077,584
;  CURRENT FILING DATE:  2002-02-15
;  PRIOR APPLICATION NUMBER: 60/269,157
;  PRIOR FILING DATE: 2001-02-15
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 1236
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(513)
US-10-077-584-3

  Query Match             1.3%;  Score 50.6;  DB 9;  Length 1236;
  Best Local Similarity   48.1%;  Pred. No. 0.28;
  Matches  143;  Conservative    0;  Mismatches  154;  Indels    0;  Gaps    0;

Qy       3220 GAATCAGCATCCACAAGTTTGAAACCAAGTATGGGTGAAAATTCTGGATTAACTACTTCT 3279
              | | || || | ||||       | ||    |     | |    | | |  |||   |  |
Db         57 GCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCA 116

Qy       3280 ACTGAAATTGAAGCTACAACAACCAGTCCTACAGAAGCTCCATCACCTGCTGTTTCTTCT 3339
              || |.||    |||  ||||| | |   |  ||| | |  || || | || |   |  |
Db        117 ACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACAGCA 176

Qy       3340 GGTACTGATGTAACTACTGAACCAACTGATACTAGAGAACAACCTACTACATTATCAACT 3399
              || |  | |||  |   | || | |  ||   |  |  | |  | |||  | ||||
Db        177 ACAACAGCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACAGCAACAACA 236

Qy       3400 ACTTCAAAAACAAACAGTGAACTGGTTGCTACTACACAAGCTACTAATGAAAATGGTGGT 3459
              |  ||| | |||  |  | |       || || |||  ||| ||      ||  |  |
Db        237 GCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCA 296

Qy       3460 AAATCTCCATCAACTGATTTAACATCAAGCTTGACAACAGGCACCTCAGCATCTACA 3516
              | | |  || || | |    | || |||      ||||||  ||  ||||| | |||
Db        297 ACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACAGCAACAACAGCAGCAACA 353


Search completed: June 20, 2003, 08:40:08
Job time : 595 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 17:11:32 ; Search time 47 Seconds
(without alignments)
3572.251 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQQFTLLFLYLSIASAKTI.....SIIQHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	938	14.4	1537	15	AAR60562	Yeast 4.7 kB agglu
2	721.5	11.1	5179	22	AAM24516	C899P predicted am
3	685	10.5	2586	22	ABB66878	Drosophila melanog
4	651	10.0	2137	23	ABP39618	Staphylococcus epi
5	640.5	9.9	2344	22	AAU37120	Staphylococcus aur
6	601	9.3	1795	22	ABB69806	Drosophila melanog
7	597	9.2	1296	23	ABG66702	Human novel polype
8	597	9.2	1296	23	ABG66756	Human novel polype
9	553.5	8.5	2870	21	AAY95559	Caenorhabditis ele
10	553.5	8.5	3178	21	AAY95556	Caenorhabditis ele
11	545.5	8.4	957	21	AAY59288	Human MUC11 polype
12	545.5	8.4	957	22	AAM24513	C900P predicted am
13	534	8.2	849	17	AAW06725	FLO1 protein, invo
14	516	7.9	894	15	AAR47578	Flocculation prote
15	516	7.9	894	15	AAR58754	S. cerevisiae FLO1
16	507	7.8	1532	21	AAB40945	Human ORFX ORF709
17	488	7.5	1721	19	AAW48299	Cryptosporidium pa
18	488	7.5	1721	21	AAB11727	Portion of Cryptos
19	488	7.5	1721	23	ABJ04045	C parvum GP900 pro
20	487	7.5	1837	21	AAB11726	Cryptosporidium pa
21	487	7.5	1837	23	ABJ04044	C parvum GP900 pro
22	468.5	7.2	862	15	AAR60563	Yeast 2.6 kB agglu
23	463	7.1	688	22	ABB30137	Peptide #2788 enco
24	463	7.1	688	22	ABB35307	Peptide #2813 enco
25	463	7.1	688	22	ABB20749	Protein #2748 enco
26	463	7.1	688	22	AAM56138	Human brain expres
27	463	7.1	688	22	AAM68511	Human bone marrow
28	463	7.1	688	22	AAM16315	Peptide #2749 enco
29	463	7.1	688	22	AAM28810	Peptide #2847 enco
30	463	7.1	688	22	AAM04053	Peptide #2735 enco
31	463	7.1	688	23	ABG38092	Human peptide enco
32	450.5	6.9	4498	22	ABB58595	Drosophila melanog
33	435.5	6.7	1045	22	ABG08332	Novel human diagno
34	428	6.6	1237	21	AAY81609	Streptococcus pneu
35	417.5	6.4	746	22	ABB59201	Drosophila melanog
36	417	6.4	560	22	AAU37464	Staphylococcus aur
37	404.5	6.2	2035	15	AAR57141	Host cell factor p
38	404	6.2	2570	22	ABG06375	Novel human diagno
39	402	6.2	2112	22	ABB60403	Drosophila melanog
40	392	6.0	502	22	AAU34408	Staphylococcus aur
41	392	6.0	1638	20	AAY00138	Enterococcus faeca
42	392	6.0	1638	20	AAY00140	Enterococcus faeca
43	392	6.0	1638	20	AAY00142	Enterococcus faeca
44	392	6.0	1638	23	ABP43357	E faecalis EF068 p
45	392	6.0	1638	23	ABP43359	E faecalis EF069 p

ALIGNMENTS

RESULT 1
AAR60562
ID AAR60562 standard; Protein; 1537 AA.
XX
AC AAR60562;
XX
DT 13-APR-1995 (first entry)
XX
DE Yeast 4.7 kB agglutination gene FLO1L.
XX
KW Yeast; agglutination; FLO1L.
XX
OS Saccharomyces cervisiae ABXL-1D.
XX
PN WO9419475-A.
XX
PD 01-SEP-1994.
XX
PF 24-FEB-1994; 94WO-JP00290.
XX
PR 26-FEB-1993; 93JP-0038871.
XX
PA (PANI-) PANIMOLABORATORIO BRYGGERILABORATORIUM.
PA (SAPB) SAPPORO BREWERIES.
XX
PI Keraenen S, Ogawa M, Onnela M, Penttila M, Takata Y;
PI Watari J;
XX
DR WPI; 1994-294338/36.
DR N-PSDB; AAQ71390.
XX
PT New yeast agglutination genes and yeast contg. them - impart
PT agglutination properties to facilitate removal from fermentation

```
PT   media
XX
PS   Disclosure; Page 43-48; 75pp; English.
XX
CC   The agglutination gene is called FLO1. Saccharomyces cerevisiae
CC   includes an agglutination gene of 4.7 kb (FLO1L) and an
CC   agglutination gene of 2.6 kb (FLO1S). FLO1L is the intact FLO1 gene
CC   on chromosome I, and FLO1S is the FLO1L gene with a portion of the
CC   ORF deleted in frame. FLO1L imparts a relatively strong
CC   agglutinative property to the host yeast into which it is introduced,
CC   while FLO1S imparts a weaker agglutinative property.
XX
SQ   Sequence   1537 AA;

  Query Match             14.4%;  Score 938;  DB 15;  Length 1537;
  Best Local Similarity   26.0%;  Pred. No. 3.6e-40;
  Matches  417;  Conservative  244;  Mismatches  523;  Indels  418;  Gaps   85;

Qy          2 LQQFTLLFLYLSIASAKT---------ITGVFDSF------NSLTWSNAANYAF------ 40
              |  |||| |  |:||  |          :|:  :|      :| |:|||| |:
Db         10 LAVFTLLAL-TSVASGATEACLPAGQRKSGMNINFYQYSLKDSSTYSNAAYMAYGYASKT 68

Qy         41 ----------------------------KGPGYPTW-----------NAVLGWSLD--G 58
                                          :   |  |            :  || |  |
Db         69 KLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGACSNSQGIAYWSTDLFG 128

Qy         59 TSANPGDTFTLNM----------PCVFKYTTSQ----------------------TSVD 85
                  | :  || |           ||: |                        || :
Db        129 FYTTPTNV-TLEMTGYFLPPQTGSYTFKFATVDDSAILSVGGATAFNCCAQQQPPITSTN 187

Qy         86 LTADGVK----------YATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGT----VTLP 131
                | ||:|          |   |:|  :          : |:  ::||    ||||
Db        188 FTIDGIKPWGGSLPPNIEGTVYMYAGYYYPM---------KVVYSNAVSWGTLPISVTLP 238

Qy        132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV-DPSAYLYASRVM 190
                   ||  | | |            :|:|       |:     || ||| |   |
Db        239 -----DGTTVSDDFEG--------YVYSFDD-------DLSQSNCTVPDPSNYA-VSTTT 277

Qy        191 PSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESF 250
               :   |  | :  |   :|| |  :   : |       |      ||   :| |
Db        278 TTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTTASTIITTTEPWNSTFTSTST 337

Qy        251 SYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTL 310
               |   :||::   :    |   :||   |  : : :|   |
Db        338 ELTTVTGTNGVRTDETII------VIRTPTTATTAITTTEPWNSTFTSTSTEL------- 384

Qy        311 RWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVTTL-PFNPSVDKTKT---------- 355
                    :  |:||:     :   || | :|||:||  |:| :   | |
Db        385 ------TTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGL 438

Qy        356 ------IEILQPIPTTTITT-------SYVGVTTSYLTKTA----PIGETATVIVDVPYH 398
                   | |  |   ||  |        ::   :|   | |     |  ||  :::  |
Db        439 PTDETIIVIRTPTTATTAMTTTQPWDDTFTSTSTEMTTVTGTNGLPTDET-IIVIRTPTT 497

Qy        399 TTT--TVTSEWTGTIT------TTTTRTN--PTDSIDTVVVQVPLPNPT-VSTTEYWSQS 447
               ||  | |  |  | |       || | ||  |||    :|:: |    | ::||: |: :
Db        498 ATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDE-TIIVIRTPTTATTAMTTTQPWNDT 556

Qy        448 FATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTVTAPP 494
               | :|    ||||   |  ||   ::|| |   |  :|||| |: :| :|    ||||
Db        557 FTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTEPWNSTFTSTSTELTTVTGTN 616

Qy        495 G-GTDS--VIIREPPNPT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPP 542
               |  ||  ::|| |   | :|||: |: :| :|    ||||   |  ||   ::|| |
Db        617 GLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPT 676

Qy        543 NPT--VTTTEYWSQSYATT----TTVTAPPG-GTDS--VIIREPPNHT--VTTTEYWSQS 591
                 |  :|||: |: :: :|    ||||   |  ||   ::|| |   |  :|||: |: :
Db        677 TATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDT 736

Qy        592 YATT----TTVTAPPG-GTD--TVIIREPPNHTV--TTTEYWSQSFATT----TTVT--- 635
               : :|    ||||   |  ||   ::|| | :  :. ||||  |: :| :|    ||||
Db        737 FTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTVTGTN 796

Qy        636 GPPSGTDTVIIREPPNP--TVTTTEYWSQSYATT----TTITAP---PGETDTVLIREPP 686
               | |:    ::|| | :     |||| |: :: :|    ||||     | :   ::|| |
Db        797 GQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMTTITGTNGVPTDETVIVIRTPT 856

Qy        687 NHTV--TTTEYWSQSYATT----TTVTAPPGE-TD--TVLIREPPNHTV--TTTEYWSQS 735
               :  :  |||| |: :: :|    ||:|    |: ||   ::|| | :  :  |||| |: :
Db        857 SEGLISTTTEPWTGTFTSTSTEMTTITGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGT 916

Qy        736 YATT----TTVTAPPG-GTD--TVIIREPPNPTV--TTTEYWSQSFATTTTVTAPPGGT- 785
               : :|    ||||   |  ||   ::|| | :  :  |||| |: :| :|:|      ||
Db        917 FTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEVTTITGTN 976

Qy        786 -----DTVIIYESMSSSKI-----------STSSNDITSIIPSFSRP------------ 816
                    :|||:  : :|  :            :::| ::|::  :  :|
Db        977 GQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSAEMTTVTGTNGQPTDETVIVIRTPT 1036

Qy        817 --HYVNSTTSD-LSTFESSSMNTPTSISSDGMLLSSTTLVTESETT----TELICSDGKE 869
                  | :||     || |:|    |   ::|:    | :| :: ||    :    | |:
Db       1037 SEGLVTTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSSSGQI 1096

Qy        870 CSRLSSSSGIVTNP--DSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVT 927
               | ::||  |:| |   || :|:::|:| ::|  |   :|:  ||::   : : :  |:
Db       1097 TSSITSSRPIIT-PFYPSNGTSVISSSVISSSVTSSLFTSSPVISSSVISSSTTTSTSIF 1155

Qy        928 TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEI 987
               :|:| :::   |   |:| :| :: ||  ||| |    :| |:  ::| |: | |   :
Db       1156 SESSKSSV--IPTSSSTSGSSESETSSAGSVSSS----SFISSESSKSPTYSSSS----L 1205

Qy        988 EIVTTSST-----KVLPPVVSSNTDLTSEPTNTREQPTTLSTTS-------NSITEDITT 1035
               :||:::|       |||  :       | | || | :: ||       ||:  | :
Db       1206 PLVTSATTSQETASSLPPATT---------TKTSEQTTLVTVTSCESHVCTESISPAIVS 1256

Qy       1036 SQPTGDNGDNTSSTN--PVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTS 1093
               :    :|  | |   |: |  |:    : |     : ::   ::     :   | |
Db       1257 TATVTVSGVTTEYTTWCPISTTETTKQTKGTTEQTTETTKQTTVVTISSCESDVCSKTAS 1316

Qy       1094 TEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTT-------SKTNSELVATT- 1145
                | :|:|:     :   ::   ::|  |::|:| | :: |       |:| |  : :|
Db       1317 PAIVSTSTATINGVTTEYTTWCPIST--TESRQQTTLVTVTSCESGVCSETASPAIVSTA 1374

Qy       1146 ----------------QATNENGGKSPSTDLTSSLTTGT-SASTSANSELVTSGSVTGGA 1188
                             |  ||     |    || | :| |: |:    : : || |
Db       1375 TATVNDVVTVYPTWRPQTANEESVSSKMNSATGETTTNTLAAETTTNTVAAETITNTGAA 1434

Qy       1189 -------------------VASASNDQSHSTSV-----TNSNSIVSNTPQTTLSQQVTSS 1224
                                  |||::   ||:||     | :   ::::  :|:|||  |:
Db       1435 ETKTVVTSSLSRSNHAETQTASATDVIGHSSSVVSVSETGNTKSLTSSGLSTMSQQPRST 1494

Qy       1225 SP------STNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
                    || :   ||| || : :   : |   |  | | |
Db       1495 PASSMVGYSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1536


RESULT 2
AAM24516
ID   AAM24516 standard; Protein; 5179 AA.
XX
AC   AAM24516;
XX
DT   12-OCT-2001  (first entry)
XX
DE   C899P predicted amino acid sequence.
XX
KW   Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW   immunogenic; gene therapy; vaccine; colonic cancer.
XX
OS   Homo sapiens.
XX
PN   WO200149716-A2.
XX
PD   12-JUL-2001.
XX
```

PF 29-DEC-2000; 2000WO-US35596.
XX
PR 30-DEC-1999; 99US-0476296.
PR 10-JAN-2000; 2000US-0480321.
PR 15-FEB-2000; 2000US-0504629.
PR 06-MAR-2000; 2000US-0519444.
PR 19-MAY-2000; 2000US-0575251.
PR 29-JUN-2000; 2000US-0609448.
PR 28-AUG-2000; 2000US-0649811.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
DR WPI; 2001-441847/47.
XX
PT Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
XX
PS Claim 2; Page 446-462; 472pp; English.
XX
CC The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.
XX
SQ Sequence 5179 AA;

Query Match 11.1%; Score 721.5; DB 22; Length 5179;
Best Local Similarity 25.3%; Pred. No. 2.5e-28;
Matches 335; Conservative 158; Mismatches 554; Indels 279; Gaps 54;

```
Qy        59 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 118
             |:  |  | | : |     ||:  |  :|      |              | | :| | :
Db      1647 TTTTPPPTTTPSPP-----TTTPSPPITTTTTPPPTT--------TPSSPITTTPSPPT 1693

Qy       119 KSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV 178
              :      | |         | |||         |  | : |        ::       :|:
Db      1694 TTMTTPSPTTTPSSPITTTTTPSSTTTPSPPPTTMTTPSPTTTPSPPTTTM------TTL 1747

Qy       179 DP---SAYLYASRVMPSLNKVT---------TLFVAPQCE-NGY-TSGTMGFSSSNGDVA 224
               |   |: |  : : ||:   |          |    | |   |:  ||   |    ||
Db      1748 PPTTTSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVPLCNWTGWLDSGKPNFHKPGGDTE 1807

Qy       225 IDCSNIHIGITKGLNDWNYPVSSESFSY----------TKTC-TSNGIQIKYQN-VPAGY 272
              :       ||   |   |   :|  :  |          |  |  | |:  | ::  | |
Db      1808 L------IGDVCG-PGWAANISCRATMYPDVPIGQLGQTVVCDVSVGLICKNEDQKPGGV 1860

Qy       273 RPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTR 332
              |             : | |  :  |      ::| |:  |  :|    :   :   ||
Db      1861 IPM----------AFCLNYEINVQCC--ECVTQPTTMTTTTTENPTPPTTTPITTTTTV 1907

Qy       333 TVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIG-ETATV 391
             | | : |   |    |    |        |  | | ||: :  ||:      | | :| |
Db      1908 TPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTT 1967

Qy       392 IVDVPYHTTTTVTSEWTGT---------ITTTT----------TRTNPTDSIDTVVVQVP 432
                 |  ||||||   | |          |||||          |:|  |  | |    |
Db      1968 ---TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTP 2024

Qy       433 LPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTA 492
             | || : |   :    ||||||  |  | |   : |    :|||        || | |
Db      2025 TPTPTGTQTP-TTTPITTTTTVTPTPTPTGT---QTPTTTPITTT--------TTVTPTP 2072

Qy       493 PPGGTD-------SVIIREPPNPTVTTTEYWSQSFATTTTVTAP---PGGTD-------S 535
             | ||         :      | || | |:  : :  ||||   |   | ||        :
Db      2073 TPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPIT 2132

Qy       536 VIIREPPNPTVTTTEYWSQSYATTTTVTAP---PGGTDSVIIREPPNHTVTTTEYWSQSY 592
                 | || | |:  : :  ||||   |   | || :     |    :|||
Db      2133 TTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-----PTTTPITTT------- 2180

Qy       593 ATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTD---------T 643
             || | |  | || |     |    :|||        || | |  |:||           |
Db      2181 -TTVTPTPTPTGTQT-----PTTTPITTT--------TTVTPTPTPTGTQTPTTTPITTT 2226

Qy       644 VIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATT 703
              :  | || | |   :    ||||:|  |  | |          | |||         ||
Db      2227 TTVTPTPTPTGTQTP-TTTPITTTTTVTPTPTPTGT--------QTPTTT-----PITTT 2272

Qy       704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTD---------TVII 754
             ||||  |  | |   : |    :|||        || | |  | ||          |  :
Db      2273 TTVTPTPTPTGT---QTPTTTPITTT--------TTVTPTPTPTGTQTPTTTPITTTTTV 2321

Qy       755 REPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFS 814
             | || | |   :    ||||||  |  | |    :: ::: |:|:    |:: |: : :
Db      2322 TPTPTPTGTQTP-TTTPITTTTTVTPTPTPTGT----QTPTTTPITTT----TTVTPTPT 2372

Qy       815 RPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLS 874
                 ||: ::|   ::  |||  :  |    :|| :| : | |      | :
Db      2373 PTGTQTPTTTPITT---TTTVTPTP-TPTGTQTPTTTPITTTTTVTPTPTPTGTQ----- 2423

Qy       875 SSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATS--SDNVSKSGVSVTTETSV 932
                 | :   :  |:  || :       :|  |: |:  :   : :|    | | :
Db      2424 -------TPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 2476

Qy       933 TTIQT---------TPNPLSSSVTSLTQLSSIPSVSESESKVT---FTSNGDNQSGTHDS 980
             ||  |         |  | :: :|: | ::  |: : :::  |    |:         : |
Db      2477 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTG 2536

Qy       981 QSTSTEIEIVTTSSTKVLP----------PVVSSNT-DLTSEPTNTREQPTTLSTTSNS 1028
               | |   | ||::    |           |: :: |   |  || |:   ||  ||: :
Db      2537 TQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTT 2596

Qy      1029 ITEDITTSQPTGDNGDNTSSTNPVPTVATST-LASASEEDNKSGSHESASTSLKPSMGEN 1087
             :|    |  ||   |  | :| |: |  | |   : :     : :  : :|:: |:
Db      2597 VT---PTPTPT---GTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPT 2650

Qy      1088 SGLTTSTEIEATTTSPTEAPSPA---------VSSGTDVTTEPTDTREQ---------- 1127
               | :|    |||: |  |:|          ::: | ||  || | |
Db      2651 GTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTT 2710

Qy      1128 -----PTTLSTTSKTNSELVATTQAT---NENGGKSPSTDLTSSLTTGTSASTSANSELV 1179
                   ||   | :| :  :  |      | ::|:|   :: || |   |   ::
Db      2711 TVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTP 2770

Qy      1180 TSGSVTGGAVA----SASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIAST 1235
             |:  :|         : :  |: :|:   : : |: ||  | :|  |::  :| | :  |
Db      2771 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPT 2830

Qy      1236 YDGSGS 1241
               :|:
Db      2831 PTPTGT 2836
```

RESULT 3
ABB66878
ID ABB66878 standard; Protein; 2586 AA.
XX
AC ABB66878;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 27426.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL10981.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 27426; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2586 AA;

Query Match 10.5%; Score 685; DB 22; Length 2586;
Best Local Similarity 24.7%; Pred. No. 8.2e-27;
Matches 405; Conservative 183; Mismatches 557; Indels 494; Gaps 67;

```
Qy         15 ASAKTITGVF-DSFNSLTWSNAANYAFKGPGYPTWNAVLGWSL--DGTSANPGDTFTLNM 71
              ||  ||| :  |  :| | |::|         |  :   | |   || |     | |
Db        102 ASTNTITKIIDDGLSSQTTSSSA---------PVVDVTQGSSSNGDGNSTQSSTTTT--- 149

Qy         72 PCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL---KSSIKAFGTV 128
                 ||: || |               | |||| |     |:       || ::  :
Db        150 ------TTTTTSSD--------------GGEFTTSSDPVVEVSQGTNGGNSSTQSSSST 188

Qy        129 TLPIAFNVGGTGSSTD---------------LEDSKCFTAGTNTVTFNDGDKDISID-- 170
                 |   : : | | ||:|                   |   |  | | : : |    | |
Db        189 TTTTSSDEGQTTSSSDPVVEVAQGSSSNGDGNSTQSLTTTTTTTTTSSDGGQSTTSSDPV 248

Qy        171 VEFEKSTVDPSAYLYAS----RVMPSLNKVTTLFVAPQCE-------NG--------YTS 211
              ||  : |   ::   :|        |    ||    |  |        ||         |:
Db        249 VEVSQGTNGGNSSTQSSSSTTTTTSSDEGQTTSSSDPVVEVAQGSSSNGDGNSTQSSTTT 308

Qy        212 GTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAG 271
                   :||:|  :   |:  : :::| |  |    | | : | | :  | |    : | |

Db        309 TTTTTTSSDGGQSTTLSDPVVEVSQGTNGGNSSTQSSSSTTTTTSSDEG-QTTSSSDPVG 367

Qy        272 YRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAG-----SNGIV 326
                              :|  :     |:  ||   |   |   :|| |     |: :|
Db        368 -----------------EVAQGSSSNGDGNSTQSST-TTTTTTTSSDGGQSTTSSDPVV 409

Qy        327 IVA------TTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVG------- 373
                |:       : | : | |  ||     | |: :|    |:      :| :|
Db        410 EVSQGTNGGNSSTQSSSATTTTT-----SSDEGQTTSSSDPVSEVAQGSSSIGDGNSTQS 464

Qy        374 -VTTSYLTKTAPIGETATV----IVDVPYHTTTTVTSEWTGTITTTTTRTN---PTDSID 425
                ||:  | |:  |  :|     :|:|   |    :|  : : ||||| ::    | | |
Db        465 STTTTTTTTTSSDGGQSTTSSDPVVEVSQGTNGGNSSTQSSSSTTTTTSSDEGQTTSSSD 524

Qy        426 TV--VVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGG-----TDTVI-----------IR 467
                |  | |    |   ::|:  | :  ||||  |:  ||     :| |:
Db        525 PVVEVAQGSSSNGDGNSTQ--SSTTTTTTTTTSSDGGQSTTSSDPVVEASQGTNGGNSST 582

Qy        468 EPPNHTVTTT----------------------------EYWSQSFATTTTVTAPPGG--- 496
                :  : | |||                          | :  |||| |:  ||
Db        583 QSSSSTTTTTSSDEGQTTSSSDPVSEVAQGSSSIGDGNSTQSSTTTTTTTTTSSDGGQST 642

Qy        497 TDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWS--- 553
                | |  : |    |          | :|||| ::  | | |        :| |   :  |
Db        643 TSSDPVVEASQGTNGGNSSTQSSSSTTTTTSSDEGQTTS-----SSDPVVEVAQGSSSNG 697

Qy        554 ------QSYATTTTVTAPPGGTDSVIIREP-----------------PNHTVTTT----- 585
                     |  |||| |    | |  : :|                  : | |||
Db        698 DGNSTQSSTTTTTTTTTSSDGGQSTTLSDPVVEGSQGTNGGNSSTQSSSSTTTTTSSDEG 757

Qy        586 ----------------------EYWSQSYATTTTVTAPPGGTDT------VIIREPPNH 616
                                   | :  |||| |:  ||  |      | : :  |
Db        758 PTTSSSDPVVEVAQGSSSNGDGNSTQSSTTTTTTTTTSSDGGESTTSSDPVVEVSQGTNG 817

Qy        617 TVTTTEYWSQSFATTTT-----------------VTGPPSGTDTVIIREPPNPTVTTTEY 659
                ::|:  | |  ||||                 |  |  |    :     | |||:
Db        818 DNSSTQ--SSSSTTTTTSSDEGQTTSSSDPVSEVAQGSSSTGDGNSTQSSTTTTTTTTDI 875

Qy        660 W---SQSYATTTTITAPPGETDT-----VLIREPPNHTVTTTEYWSQSYATTTTVTAPPG 711
                :    |   |||| ::  ||: |     | : :  |   ::|:    | :|||| ::  |
Db        876 FRCPRQRTTTTTTTSSDGGESTTLSDPVVEVSQGTNGGNSSTQ---SSSSTTTTSSDEG 932

Qy        712 ETDT-----VLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDT------VIIREPPNP 760
                :| :     | : :  :          | :  |||| |:  ||  |      | : :  |
Db        933 QTTSSSDPVVEVAQGSSSNGDGNSTQSSTTTTTTTTTSSDGGESTTSSDPVVEVSQGTNG 992

Qy        761 TVTTTEYWSQSFATTTTVTAPPGGTDT-----VIIYESMSSSK---ISTSSNDITSIIPS 812
                ::|    || ::|||  |:   |  |     |:     |||     || |:  |:|  :
Db        993 DNSST----QSSSSTTTTTSSDEGQATSSSAPVVDISQGSSSNGDGNSTQSSTTTTITTT 1048

Qy        813 FSRPHYVNSTTSD-----------LSTFESSSMNTPTSISSDGMLLSSTTLVTESETTT 860
                |     ::|:||              |: :|||  | |: | :|   ||      | :
Db       1049 TSSDGDQSTTSSDPVVEVSQGTNGGNSSTQSSSSTTTTTSSDEGQTTSSFGSCRRSHSRI 1108

Qy        861 ELIC-----------------------------SDGK--ECSRLSSSSGIVTNPDSNESSI 890
                 |                            | |   : |   |||    |   |:|
Db       1109 LFQCPRQRQQLQRRHLPMVANPPPHLTPVVEVSQGTNGDNSSTQSSSSTTTTTSSDEGQT 1168

Qy        891 VTSTVPTASTMSDSLSST--DG-ISATSSDNV-------------SKSGVSVTTETSVTT 934
                :|: | :|| : : ::|  ||  | |||| |             ::|  | || ||
Db       1169 TSSSAPVSSTTTTTTTTTSSDGDQSTTSSDPVVEVSQGTNGDNSSTQSSSSTTTTTSSDE 1228

Qy        935 IQTTP-----------------NPLSSSVTSLTQLSSIPSVSESES------KVTFTSN 970
                |||                 |   || |: |  ::     || :      :|:  :|
Db       1229 GQTTSSSAPVVEVTQGSSSNGDGNSTQSSTTTTTTTTTSSDGGESTTSSDPVVEVSQGTN 1288

Qy        971 GDNQSGTHDSQSTSTEIEIVTTSS-----TKVLPPVV------SSNTDLTSEPTNTREQP 1019
                ||| | |  | ||:|     ||||     |   |||      ||| |  |  ::|
Db       1289 GDN-SSTQSSSSTTT-----TTSSDEGQTTSSSAPVVEVTQGSSSNGDGNSTQSST--TT 1340

Qy       1020 TTLSTTSNSITEDITTSQPT-----GDNGDNTSSTNPVPTVATSTLASASEEDNKSG--- 1071
                || :|||:   |  |:| |      | |||| |||    :  |:| :   :  : ||
Db       1341 TTTTTTSSDGGESTTSSDPVVEVSQGTNGDN-SSTQSSSSTTTTTSSDEGQTTSSSGPVV 1399
```

```
Qy       1072 --SHESAS-----------------------TSLKPSMGENSGLT-----------TS 1093
Db       1400 DISQGSSSNVLDNDNNYNDDIFRWWANPPPYLTPLWKSVKEQMAITAPLSLRRPPQQQTS 1459

Qy       1094 TEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE----------LVA 1143
Db       1460 SD-EGQTTS-SSAPVVDISQGSSSNGDGNST-QSSTTTTTTTTTSSDGGESTTLSDPVVE 1516

Qy       1144 TTQATNENGGKSPSTDLTSSLTTGTSA----STSANSELV--------------TSGSVT 1185
Db       1517 VSQGTN---GDNNSTQSSSSTTTTTSSDEGQTTSSSDPVVEVAQGSSSNGDGNSTQSSTT 1573

Qy       1186 GGAVASASNDQSHST----------------SVTNSNSIVSNTPQTTLSQQVTSSSPST 1228
Db       1574 TTTTTTTSSDGGESTTSSDPVVEVSQGTNGDNSSTQSSSSTTTTTSSDEGQTTSSSDPVV 1633

Qy       1229 NTFIASTYDGSGSIIQHST 1247
Db       1634 EVAQGSSSNGDGNSTQSST 1652
```

RESULT 4
ABP39618
ID ABP39618 standard; Protein; 2137 AA.
XX
AC ABP39618;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4463.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2002-381255/41.
DR N-PSDB; ABN92163.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS Disclosure; SEQ ID 4463; 267pp; English.
XX
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 2137 AA;

Query Match 10.0%; Score 651; DB 23; Length 2137;

Best Local Similarity 24.2%; Pred. No. 3.7e-25;
Matches 314; Conservative 252; Mismatches 541; Indels 192; Gaps 37;

```
Qy         19 TITGV-----FD-SFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMP 72
Db        745 TVTGLPQGLKFDASTNSIV-----------GTPTQIGTNTITIESTDASGNK------- 785

Qy         73 CVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPI 132
Db        786 -----TTTKINYEVTRNSASDSTSTSIVNSVSTSISNST-----SLSDSVKASQSL---- 831

Qy        133 AFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEK--STVDPSAYLYASRVM 190
Db        832 ---------STSKSLSESLSASTSNSTSIQASESASTSKQLSESASTSTSDSASESARKS 882

Qy        191 PSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESF 250
Db        883 ESTSKSTSL-----SESTSTSVSDSASVSTSESASTSTSVSGSTSTSISD---------- 927

Qy        251 SYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFT- 309
Db        928 -STSTSTSDSASIKASESASTSKLLSESVSTSTSDSASTSTSVSDSNSASTSLSKSTSTS 986

Qy        310 -------------LRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTI 356
Db        987 VSDSTSTSTSDSASTSTSESESDSASTSL-SESTSTSVSDSTSTSTSDSASMSASESESN 1045

Qy        357 EILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTT 416
Db       1046 SKSTSLSEST-STSLSGSTSASTSDSA-----STSTSESESDSTSTSLSESTSTSLSGST 1099

Qy        417 RTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTT 476
Db       1100 SASTSDSAST-----------STSESDSTSESTSLSESLSTSVSDSTSASTSESASTST 1147

Qy        477 TEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSV 536
Db       1148 SESESNSASTSLSGSLSTSISDSTSTSTSDSASTSTSESESDSTSTSLSESTSTSLSDST 1207

Qy        537 IIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTT 596
Db       1208 STSTSESASTSTSESDSTSESTSLSESTSTSVSDSTSASTSDSASTSTSVSDSESASTSI 1267

Qy        597 TVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTT 656
Db       1268 SESLSTSVSDSTSTSTSDSASTSTSESDSTSESTSLSESISTSVSDSTSASTSDSASTST 1327

Qy        657 TEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTT-------VTAP 709
Db       1328 SE--SESDSASTSLS---GSTSTSL----SDSTSTST---SDSASTSTSESDSERASTSL 1375

Qy        710 PGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWS 769
Db       1376 SGSTSTSL----SDSTSTST---SDSASTSTSVSDSNSASTSL------SGSLSTSVSDS 1422

Qy        770 QSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTF 829
Db       1423 TSTSTSDSASASTSESDSERASTSLSGS-TSTSISDSTSTSTS-------DSASTSTSVS 1474

Qy        830 ESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESS 889
Db       1475 ESNSTSTSISESLSTSVSDSTSTSTSDSASTSTSVSDSDSAS--TSSSESVSTSDSESTS 1532

Qy        890 IVTS-TVPTASTMSDSLSSTDGISATSSDNVSKS-GVSVTTETSVTTIQTTPN--PLSSS 945
Db       1533 TSTSDSASTSTSVSESNSTSTSLSGSTSTSVSDSTSTSTSDSASASTSESDSDSASTSSS 1592

Qy        946 VTSLTQLSSIPSVSESESKVTFTSNGD-NQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSS 1004
Db       1593 ESVSTSVSDSTSASTSESASTSTSVSDSNSASTSLSESTSTSLSDSTSMSTSDSASTSTS 1652

Qy       1005 NTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASAS 1064
```

```
Db       1653 ESDSDSASTSLSD------STSTSVSESTSTSTSTSVSASNSTSTSLSDSRSTSLSDSTS 1706

Qy       1065 EEDNKSGS-------HESASTSLKPSMGE--NSGLTTSTEIEATT----------TSPT 1104
                ::|||        :||||||  |     :   :|||   |:|           || :
Db       1707 TSTSESGSTSTSESDSDSASTSLSESTSTSISDSTSTSTSDSASTSMSVSDSNRASTSLS 1766

Qy       1105 EAPSPAVSSGTDVTTE---PTDTREQPTTLSTTSKTNSELVATTQ----------ATN 1149
              :: | :||  |  :|    |||||  :| ::|| : |   : :             :|:
Db       1767 DSTSTSVSDSTSASTSESASTSTRESESTSASTSLSESTSTSVSDSTSTSTSDSASTSTS 1826

Qy       1150 ENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIV 1209
              |:   | || |: | :|  | ||||:    || | :     | ||  | | | : | |:
Db       1827 ESDSNSESTSLSESTSTSVSDSTSAS----TSASASTSTSVSDSNSASTSLSGSTSTSVS 1882

Qy       1210 SNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSI-IQHST 1247
                :|  :| :   ||:| | :   :::  || |  |  ||
Db       1883 DSTSTSTSASASTSTSESDSDSASTSLSGSTSTSISDST 1921
```

RESULT 5
AAU37120
ID AAU37120 standard; Protein; 2344 AA.
XX
AC AAU37120;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #1290.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS54979.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 12713; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2344 AA;

Query Match 9.9%; Score 640.5; DB 22; Length 2344;
Best Local Similarity 22.2%; Pred. No. 1.4e-24;
Matches 323; Conservative 269; Mismatches 542; Indels 323; Gaps 46;

```
Qy         17 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 76
              | |:    ||  |  :|   |   |  |:      :|::      :||
Db        297 AITLGTRIDSNKSFHFSGKVNLGNKYEGHGNGGDGIGFAF-----SPG------------ 339

Qy         77 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIK--------AFGTV 128
                :| ::  | |:   :  |  | :  |:   | | | : |:           |||
Db        340 -VLGETGLNGAAVGIGGLSNAF--GFKLDTYHN-TSTPNSSAKAKADPSNVAGGGAFGAF 395

Qy        129 TLPIAFNVGGTGSSTDLED--SKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYA 186
                :: |  | :|:   |  :|     ||  || |       |: :   |
Db        396 VTTDSYGVASTYTSSSTADNAAKLNVQPTNN-TFQD------FDINYNGDT--------- 439

Qy        187 SRVMPSLNKVTTLFVAPQ----------CENGYTSGTMGFSSSNGDVAIDCSNIHIGITK 236
                  || |:  | |              ::| |: ::  ::| |  | :   :  |
Db        440 --------KVMTVKYAGQTWTRNISDWIAKSGTTNFSLSMTASTGG-ATNLQQVQFG--- 487

Qy        237 GLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAG--------YRPFIDAYISATDVNQY 288
                        :| ||::  :   |::| :|  |         |   :|  ::  :
Db        488 ------------TFEYTESAVT---QVRYVDVTTGKDIIPPKTYSGNVDQVVTIDNQQSA 532

Qy        289 TLAYTNDYTCAGSRLQS---------------KPFTLRWT---------GYKNSDAGS-- 322
                |   :||   |  |                :  |  :|         | :  : |
Db        533 LTAKGYNYTSVDSSYASTYNDTNKTVKMTNAGQSVTYYFTDVKAPTVTVGNQTIEVGKTM 592

Qy        323 NGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPI------------------PT 364
              | ||:  |       |  || ||   | |      | |                    |
Db        593 NPIVLTTTDNGTGTVTNTVTGLPSGLSYDSATNSIIGTPTKIGQSTVTVVSTDQANNKST 652

Qy        365 TTITTSYVGVTTSYLTKTAPIGETATVI-------------------------------- 392
              || | : |  |   :|   |||: :: :
Db        653 TTFTINVVDTTAPTVT---PIGDKSSEVFSPISPININIATQDNSGNAVTNTVTGLPSGLTF 709

Qy        393 -----------VDVPYHTTTTVTSEWTGTITTTT-----TRTNPTDSIDTVVVQVPLPNP 436
                        ::   | | |::: :|  ||||     || : :||: |        :
Db        710 DSTNNTISGTPTNIGTSTITIVSTDASGNKTTTTFKYEVTRNSMSDSVST--SGSTQQSQ 767

Qy        437 TVSTTEYWSQSFATT------TTVTAPPGGTDTVIIRE--PPNHTVTTTEYWSQSFATTT 488
              :|||::  ||| :|:       |: :|    : :| : :    : :::|:|  | | :|:|
Db        768 SVSTSKADSQSASTSTSGSIMTSTSASTSKSTSVSLSDSVSASKSLSTSESNSVSSSTST 827

Qy        489 TVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTT 548
              ::       : |  :    : : : ::: | | :|  : :     :||:      : :|:
Db        828 SLVNSQSVSSS--MSGSVSKSTSLSDFISNSSSTEKSESVSTSTSDSLRTSTSLSDSVSM 885

Qy        549 TEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAP------- 601
              :   | | : : : :     : |  : :  :::::|:|  |:| :|: :::
Db        886 STSGSLSKSQSLSTSTSDSASTSQSVSDSTSNSISTSESLSESGSTSESISISNSISNSV 945

Qy        602 -------PGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTV 654
                      : :: :    : :::|:|  | | :|: :|:|  |   :  :    : ::
Db        946 SASTSKLESQSTSISLSTSDSKSMSTSESLSDSTSTSDSVSGSLSVAGSQSVSTSTSDSM 1005

Qy        655 TTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATT-------TTVT 707
              :|:|  | | :|: :: |   :: :|      : : :|:|  | | :|:       : |
Db       1006 STSEMISDSMSTSGSLAASDSKSMSVSSSMSTSQSGSTSESLSDSISTSDSDSKSLSLST 1065

Qy        708 APPGETDTVL-----IREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTV 762
              :  | | |       :|   : : : :   ||| :|: : :     :|:       : :
Db       1066 SQSGSTSTSTSTSSSVRMSESQSTSGSMSTSQSDSTSISTSFSDSTSDS------KSAST 1119
```

```
Qy        763 TTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNST 822
Db       1120 ASSESISQSVSTSTS-----GSVSTSTSLSTSNSERTSTSMSDSTSLSTSESDSTSDSTS 1174

Qy        823 TSDLSTFESSSMNTPTSIS-------SDGMLLSSTTLVTE------SETTTELICSDGKE 869
Db       1175 TSD-SISEAISGSESTSISLSESNSTSDSESKSASAFLSESLSESTSESTSESLSGSTSD 1233

Qy        870 CSRLS-------SSSGIVTNPDSNESSIVTSTVPTAST--------------MSDSLSS 907
Db       1234 STSLSDSNSESGSTSTSLSNSTSGSASISTSTSGSASTSTVKSESVSTSLSTSTSTSLSD 1293

Qy        908 TDGISATSSDNVSKS-----GVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESE 962
Db       1294 STSLSTSLSDSTSGSKSNSLSASMSTSDSISTRKSESLSASTSLSGSTSESESGSTSSSE 1353

Qy        963 SKVTFTS---------NGDNQSGTHD--SQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSE 1011
Db       1354 SKSDSTSMSLSMSQSTSGSTSVSTSESLSDSTSTSLSLSASMNQSGVDSNSASQSASTST 1413

Qy       1012 PTNTRE--QPTTLSTTSNSITEDITTSQPT--GDNGDNTSSTNPVPTVATSTLASASEED 1067
Db       1414 STSTSESDSQSTSSYTSQSTSQSESTSTSTSLSDSTSISKSTSQSGSTSTSASLSGSESE 1473

Qy       1068 NKSGS---------HESASTSLKPSMG-ENSGLTTSTEIEATTTSPTEAPSPAVSSGTDV 1117
Db       1474 SDSQSISTSTSESKSESTSTSLSDSTSTSNSGSASTSTLLSNSASASESDSSSTSL-SDS 1532

Qy       1118 TTEPTDTREQPTTLSTTSKTNSELVAT------------TQATNENGGKSPSTDLTSSLT 1165
Db       1533 TSASMQSSESDSQSTSTSLSNSQSTSTSIRMSTIASESVSESTSESGSTSESTSESDSTS 1592

Qy       1166 TGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSS 1225
Db       1593 TSLSDSQSTS----RSTSASGSASTSTSTSDSRSTSASTSTSMRT----STLDSQSMSLS 1644

Qy       1226 PSTNTFIASTYDGSGSI 1242
Db       1645 TSTSTSVSDSTSLSDSV 1661
```

RESULT 6
ABB69806
ID ABB69806 standard; Protein; 1795 AA.
XX
AC ABB69806;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 36210.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL13909.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 36210; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1795 AA;

Query Match 9.3%; Score 601; DB 22; Length 1795;
Best Local Similarity 24.7%; Pred. No. 1.1e-22;
Matches 323; Conservative 162; Mismatches 499; Indels 322; Gaps 57;

```
Qy        126 GTVTLPIAFNV--GGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAY 183
Db        182 GTIFSPVERKCLPGDQCPSTEISDSGSYIPQNCELKFPECAEEGTF-----RSPTDCALY 236

Qy        184 LYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNY 243
Db        237 -YTCRLQESGTYLQTRFKCP----GSNSFDLERKLCRPRSEVDCFDFVPG---------- 281

Qy        244 PVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSR- 302
Db        282 PV--------------------QVPYAPQPYYPPYPAAPPL------YEEDDYDTGARE 314

Qy        303 ----LQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVD------- 351
Db        315 QQPALKSEKLQVAAEGFEKP---SLNVVVLQT--TTLEPSTAYHKYPAYPSYPSYEYSSH 369

Qy        352 --------------------KTKTIE--ILQP-IPTTTITT--------SYVGVTTSYLT 380
Db        370 HRGKERAAENLELEKEGVPRKLKLSENIVIQPETPATAATTREPLNDINKYQYKRYTYGT 429

Qy        381 KTAPIGETATV-----------IVDVPYHTTTTVTSEWTGTITTTTTR---TNPTDSID 425
Db        430 DKNDVTEAPEIKSPLKGLHLSENIVILPETTTTTTTTTKPVVLTCPTISPPDTTPKPSTT 489

Qy        426 TVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTD-TVIIREPPNHTVTTTEYWSQSF 484
Db        490 TAVTK---STPKISSTEQHSTTTAKTTTTKRPTTVTEKTSSATEKPRTTVVTTT--TQKR 544

Qy        485 ATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNP 544
Db        545 STTTHNTSP--DTKTTIRSTTLSPKTTTTP------STTTPSTTTPSTT---------TP 587

Qy        545 TVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTT--TVTAPP 602
Db        588 STTTP-------STTTPSTTTPSTTTTVKV---STHRPRTTSQKTTTASTTTKKTTTSPK 637

Qy        603 GGTDTVIIREPPNHTVTTTE---YWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEY 659
Db        638 TTKTTDIPTSTTSKLSTTTQKTTTTTHKFTAATTSTEKPKTT-----TEKTSTVSTTTKK 692

Qy        660 WSQSYATTTTITAPPGETDTVLIREPPNHT--VTTTEYWSQS---YATTTTVTAPPGETD 714
Db        693 STESSPKPTSSTGKPTTTPKPSTRTTPTTTKVTTTTQITTTTPLRSSTETTSTQPPTTT- 751

Qy        715 TVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFAT 774
Db        752 ------TPQPTTTTTLTVTPKTSTTTTTTEKP-------ITSSPKPT-TTTQKTTSTAPN 797

Qy        775 TTTV-------TAPPGGTDTVIIYESMSSSKISTSSND--ITSIIPSFSRPHYVNSTTSD 825
```

```
Db        798 TTKVAITTQKETTPTQSTSTTIFTRKTTTNNPEPTSTEKPITSTTP---KP---STTTPK 851

Qy        826 LSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDS 885

Db        852 TSTVASST--EKTTISSP---KPTTEKSTENPTTNSV------KTSALTSSTQRATSTTS 900

Qy        886 NESSI---VTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTE-----TSVTTI-- 935

Db        901 EPTKTTQNITTTTPKPTTLKTSTQE----ATTSTQKVST--VTITTKKATESSPLTTLST 954

Qy        936 ---QTTPNPL------SSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQST--- 983

Db        955 EEPNTTPKPLRTTTPTTTSVTATTRITT-TTISESSTETTSTQKPKSTTPTSTTRTTPKV 1013

Qy        984 -----STEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT--TLSTTSNSITEDITTS 1036

Db       1014 TTVIVSTQNPTTTTSKTSTVTITTPNPSPSTQRPTTTTRQPTSITASTTSIGTTRIPTTT 1073

Qy       1037 QP----------------------TGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHE 1074

Db       1074 NPQNSTSSTDLTTVTRPPCPDPDSTSDKNTNTACTQELQQV--NLLELQSPQKQEQFTHT 1131

Qy       1075 SASTSLKPSMGENSG---------LTTSTEIEATTTSPTEAPSPAVSSGTD--------V 1117

Db       1132 RTHTALTGSRNTLGGQEVPDYMDDAPSSAEAESGQATTAKAPTMSTLAAAHLLQKLFHII 1191

Qy       1118 TTEPTDTREQPTTLSTTSKTNSEL-------------VATTQ--------------ATNE 1150

Db       1192 STTPPSREHAPTQRPSSQPSSSQRSRGVTIAQMARHNLATSKPFIAHSLRLSIQQLASTQ 1251

Qy       1151 NGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVS 1210

Db       1252 KRSIPPKTLVTHNTTKEPEDSEYYDSE--TSEQYTDEDNEVLDKTQPRAMSSTTVAAVLP 1309

Qy       1211 NTPQTTLSQ--QVTSSSPSTNTFIAST-------------YDGSGS 1241

Db       1310 AVPSTTTEREPQKTSSSPSPTKATSSTTTQPIETTTGDLEYDSSGS 1355
```

RESULT 7
ABG66702
ID ABG66702 standard; Protein; 1296 AA.
XX
AC ABG66702;
XX
DT 30-AUG-2002 (first entry)
XX
DE Human novel polypeptide #37.
XX
KW Human; inflammatory condition; shock; sepsis; immune response;
KW cancer; wound healing; central nervous system disease; haematopoiesis;
KW peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW bone degenerative disorder; periodontal disease; reperfusion injury;
KW lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW allergic condition; thrombolysis; thrombosis; coagulation disorder;
KW fungal infection.
XX
OS Homo sapiens.
XX
PN WO200244340-A2.
XX
PD 06-JUN-2002.
XX
PF 30-NOV-2001; 2001WO-US47004.
XX
PR 30-NOV-2000; 2000US-0028952.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;
PI Yamazaki V, Ujwal ML, Drmanac RT;
XX
DR WPI; 2002-508509/54.
DR N-PSDB; ABK94926.
XX
PT Novel nucleic acids and polypeptides for diagnosis, treatment of
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT disorders, cancer and promoting wound healing -
XX
PS Claim 10; Page 604-607; 672pp; English.
XX
CC The invention relates to human novel polynucleotides and associated
CC polypeptides. The polynucleotides and polypeptides are useful for
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC and cancer and for promoting wound healing. The sequences are used to
CC induce the proliferation of neural cells and regeneration of nerve and
CC brain tissue, and are useful for the treatment of central and peripheral
CC nervous system diseases and neuropathies, such as Alzheimer's disease,
CC Parkinson's disease, Huntington's disease and amyotrophic lateral
CC sclerosis. The sequences are involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC cell disorders and platelet disorders such as thrombocytopenia,
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC disease. The sequences of the invention are also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, immune deficiencies and disorders
CC including severe combined immunodeficiency (SCID), bacterial or fungal
CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC and coagulation disorders. Sequences ABG66666-ABG66758 represent human
CC novel polypeptides of the invention.
XX
SQ Sequence 1296 AA;

Query Match 9.2%; Score 597; DB 23; Length 1296;
Best Local Similarity 24.1%; Pred. No. 1.2e-22;
Matches 344; Conservative 203; Mismatches 464; Indels 414; Gaps 67;

```
Qy         12 LSIASAKTITGVFDSF----NSLTWSNAANYAFKGPGYPTWNAVLGWSLDGT------SA 61

Db         21 VSMTSA-TIPSVRPTFTSTHNTLTSSLLTTF----PGTYSFSSSMSASSDGTTHTETITS 75

Qy         62 NPGDTFTLNMPC--------VFKYTTSQT------SVDLTADGVKYATCQFYSGEEFTTF 107

Db         76 LPASTSTLHTTAESTTAHTTTTSFTTSTTMESPSSSVATTSTG---QTTFSSSTATFTET 132

Qy        108 STLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI 167

Db        133 TTLTPT-TDFSEETLTTAMTSTPPITSSITPTNTVTSMTTMTSWPTATNTLS-------- 183

Qy        168 SIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDC 227

Db        184 SLTTNILSSTPVPSTERTTSHT-TNINPVSTL--------------------------V 215

Qy        228 SNIHIGITKG--LNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDV 285

Db        216 TTLPTTITRSTPTSETTYPISS-----TSTVTESTTEITY-------------------- 250

Qy        286 NQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLP 345

Db        251 ---------------------------------------------STTMTETSSSATSLP 265

Qy        346 F-NPSVDKTKTIE----ILQPIPTTTITTSYVGVTTSYLTKTAPIGETA----TVIVDVP 396

Db        266 LTSPLVSTTETAKTPTTIL--VTTTTKTTSH--STTSFTSST--VYSTASTHTTAITSVP 319

Qy        397 YHTTTTVTSEWTGTITTTTTRTNPTDSIDTV-----VVQVPLPNPTVSTTEYWSQSFATT 451

Db        320 TTLGTMVTS--TSRIPSTVSTSIPTSQPKTVNSSSGGITGSLPMMTDLTSGY------TV 371
```

```
Qy        452 TTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPG-GTDSVIIREPPNPTV 510
              ::::| |    | :|  | : ||  ||  :  | : |: | | |  | :  :  |    :
Db        372 SSMSAIP----TTVI--PTSLTVQNTE--TSIFVSMTSATTPSGRPTFTSTVNTPTRSLL 423

Qy        511 T---TTEYWSQSFAT-------TTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTT 560
              |   ||  :| | :        | ::::||  | ::       |: |||     |: |:|
Db        424 TSFPTTHLFSSSMSESSAGTTHTESISSPPATTSTLHTTAESTPSCTTT----TSFITST 479

Qy        561 TVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTT 620
              |:             || :  | ||          ||||     : |   ||    | ||
Db        480 TM------------EPLSTIVATTG------TVKTTVT-----SSTATFRETTTLTSTT 515

Qy        621 ---TEYWSQSFATTTTVTGPPSGTDTV------IIREPPNPTVT--TTEYWSQ------- 662
                 ||    :  :|| :|   :   |:        |   | |::  |:   |
Db        516 DISTESLMTAMTSTTRLTSAITSKTTLTSLKTTASRPTANSTLSSLTSSILSSTLVPSTD 575

Qy        663 ---SYATTTTITAPPGET--DTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVL 717
                 |: |  | ::|   |   |:    | : |   |   |    :||::      | |:
Db        576 MITSHTTNLTRSSPLLATLPTTITRSTPTSETTYPTSPTSTVKGSTTSIRYSTSMTGTLS 635

Qy        718 IRE--PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATT 775
              :    ||          | |  || | |  |    | :|   ||    || : : || ::
Db        636 METSLPPT---------SSSLPTTETATMTP---TTTLITTTPN----TTSHSTPSFTSS 679

Qy        776 TTVTAPPGGTDTVIIYESMSSSKISTSSNDITS---IIPSFSRPHYVNS--TTSDLSTFE 830
              |             || ::|:|  : :|:  ||   : |:   |  :::   |: | |
Db        680 T------------IYSTVSTSTTAITSHFTTSETAVTPTPVTPSSLSTDIPTTSLRTLT 726

Qy        831 SSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSI 890
               ||: | ||:::     |  | ::   | | :|          |||  |     :  ||:
Db        727 PSSVGTSTSLTTTTDFPSIPTDISTLPTRTHII----------SSSPSI---QSTETSSL 773

Qy        891 VTSTVPTASTMSDSLSSTDG----------------------ISATSSDNVSKSGVSVT 927
              | :| || ||:  :|  |:                     ::: :    | :  :||
Db        774 VGTTSPTMSTVRMTLRITENTPISSFSTSIVVIPETPTQTPPVLTSATGTQTSPAPTTVT 833

Qy        928 ---TETSVTTIQT-TPNPLSSSVTSLTQ----------LSSIPSVSESESKVTFTSNGDN 973
                 |::| :|: | ||:   |:: | :|          | : ||    :: :| ||
Db        834 FGSTDSSTSTLHTLTPSTALSTIVSTSQVPIPSTHSSTLQTTPSTPSLQTSLTSTSEFTT 893

Qy        974 QSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTD---LTSEPTNTREQPTTLSTTSNSI- 1029
              :| |  | ||:    |:|: ||     :: |:|    ::| |::    |   ||| :|:
Db        894 ESFTRGSTSTNA---ILTSFST-----IIWSSTPTIIMSSSPSSASITP-VFSTTIHSVP 944

Qy       1030 -------TEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLK- 1081
                    ||:: ::  ||           |::::|   | |   : | |  :| | :
Db        945 SSPYIFSTENVGSASITG----------FPSLSSSATTSTS---STSSSLTTALTEITP 990

Qy       1082 ------PSMGENSGLTTSTEIEATTTSP-------TEAPSPAVSS------GTDVTTEPT 1122
                    ||     |  | | : |: | |        ||| ||  :         |:: | ||
Db        991 FSYISLPSTTPCPGTITITIVPASPTDPCVEMDPSTEATSPPTTPLTVFPFTTEMVTCPT 1050

Qy       1123 DTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDL--TSSLTTGTSASTS----ANS 1176
                | |||:|   |:| :  :  : : |   |  |    |::: | |   ||    :||
Db       1051 SISIQ-TTLTTYMDTSSMMPESESSISPNASSSTGTGTVPTNTVFTSTRLPTSETWLSNS 1109

Qy       1177 ELV----------------TSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQ 1220
               ::                :|   |    :| |   |  |: |:   : :    |||: :
Db       1110 SVIPLPLPGVSTIPLTMKPSSSLPTILRTSSKSTHPSPPTTRTSETPVATTQTPTTLTSR 1169

Qy       1221 VTS------SSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLF 1259
                |:      :: || |  | | |  |       || |    |  |
Db       1170 RTTRITSQMTTQSTLTTTAGTCDNGG------TWEQGQCACLPGF 1208
```

```
RESULT 8
ABG66756
ID   ABG66756 standard; Protein; 1296 AA.
XX
AC   ABG66756;
XX
DT   30-AUG-2002  (first entry)
XX
DE   Human novel polypeptide #91.
XX
KW   Human; inflammatory condition; shock; sepsis; immune response;
KW   cancer; wound healing; central nervous system disease; haematopoiesis;
KW   peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;
KW   myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;
KW   cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;
KW   bone degenerative disorder; periodontal disease; reperfusion injury;
KW   lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;
KW   allergic condition; thrombolysis; thrombosis; coagulation disorder;
KW   fungal infection.
XX
OS   Homo sapiens.
XX
PN   WO200244340-A2.
XX
PD   06-JUN-2002.
XX
PF   30-NOV-2001; 2001WO-US47004.
XX
PR   30-NOV-2000; 2000US-0028952.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Goodrich RW,  Liu C,  Zhou P,  Asundi V,  Wang J,  Wang D;
PI   Yamazaki V,  Ujwal ML,  Drmanac RT;
XX
DR   WPI; 2002-508509/54.
DR   N-PSDB; ABK94980.
XX
PT   Novel nucleic acids and polypeptides for diagnosis, treatment of
PT   inflammatory, autoimmune, nervous system, myeloid or lymphoid cell
PT   disorders, cancer and promoting wound healing -
XX
PS   Claim 10; Page 669-671; 672pp; English.
XX
CC   The invention relates to human novel polynucleotides and associated
CC   polypeptides. The polynucleotides and polypeptides are useful for
CC   treating inflammatory conditions such as arthritis, nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis, immune responses
CC   and cancer and for promoting wound healing. The sequences are used to
CC   induce the proliferation of neural cells and regeneration of nerve and
CC   brain tissue, and are useful for the treatment of central and peripheral
CC   nervous system diseases and neuropathies, such as Alzheimer's disease,
CC   Parkinson's disease, Huntington's disease and amyotrophic lateral
CC   sclerosis. The sequences are involved in chemotactic or chemokinetic
CC   activity, regulation of haematopoiesis, treatment of myeloid or lymphoid
CC   cell disorders and platelet disorders such as thrombocytopenia,
CC   regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
CC   growth, tissue repair, healing of burns, incisions, ulcers, treatment of
CC   osteoporosis, osteoarthritis, bone degenerative disorders and periodontal
CC   disease. The sequences of the invention are also useful for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues, immune deficiencies and disorders
CC   including severe combined immunodeficiency (SCID), bacterial or fungal
CC   infections, autoimmune disorders e.g. multiple sclerosis and myasthenia
CC   gravis, allergic conditions such as asthma, thrombolysis or thrombosis
CC   and coagulation disorders. Sequences ABG66666-ABG66758 represent human
CC   novel polypeptides of the invention.
XX
SQ   Sequence   1296 AA;

  Query Match             9.2%;  Score 597;  DB 23;  Length 1296;
  Best Local Similarity   24.1%;  Pred. No. 1.2e-22;
  Matches  344;  Conservative  203;  Mismatches  464;  Indels  414;  Gaps   67;

Qy         12 LSIASAKTITGVFDSF----NSLTWSNAANYAFKGPGYPTWNAVLGWSLDGT------SA 61
              :|: || ||  |  :|    |:|| |    :   ||  :::: :  | |||      ::
Db         21 VSMTSA-TIPSVRPTFTSTHNTLTSSLLTTF----PGTYSFSSSMSASSDGTTHTETITS 75

Qy         62 NPGDTFTLNMPC--------VFKYTTSQT------SVDLTADGVKYATCQFYSGEEFTTF 107
                 | | ||:           :||| |       ||  |: |    |   |   ||
Db         76 LPASTSTLHTTAESTTAHTTTSFTTSTTMESPSSSVATTSTG---QTTFSSSTATFTET 132
```

```
Qy       108 STLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDI 167
Db       133 TTLTPT-TDFSEETLTTAMTSTPPITSSITPTNTVTSMTTMTSWPTATNTLS-------- 183

Qy       168 SIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDC 227
Db       184 SLTTNILSSTPVPSTERTTSHT-TNINPVSTL--------------------------V 215

Qy       228 SNIHIGITKG--LNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDV 285
Db       216 TTLPTTITRSTPTSETTYPISS-----TSTVTESTTEITY---------------------- 250

Qy       286 NQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLP 345
Db       251 ---------------------------------------------STTMTETSSSATSLP 265

Qy       346 F-NPSVDKTKTIE----ILQPIPTTTITTSYVGVTTSYLTKTAPIGETA----TVIVDVP 396
Db       266 LTSPLVSTTETAKTPTTIL--VTTTTKTTSH--STTSFTSST--VYSTASTHTTAITSVP 319

Qy       397 YHTTTTVTSEWTGTITTTTTRTNPTDSIDTV-----VVQVPLPNPTVSTTEYWSQSFATT 451
Db       320 TTLGTMVTS--TSRIPSTVSTSIPTSQPKTVNSSSGGITGSLPMMTDLTSGY------TV 371

Qy       452 TTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPG-GTDSVIIREPPNPTV 510
Db       372 SSMSAIP----TTVI--PTSLTVQNTE--TSIFVSMTSATTPSGRPTFTSTVNTPTRSLL 423

Qy       511 T---TTEYWSQSFAT-------TTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTT 560
Db       424 TSFPTTHLFSSSMSESSAGTTHTESISSPPATTSTLHTTAESTPSCTTT----TSFITST 479

Qy       561 TVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTT 620
Db       480 TM------------EPLSTIVATTG------TVKTTVT-----SSTATFRETTTLTSTT 515

Qy       621 ---TEYWSQSFATTTTVTGPPSGTDTV------IIREPPNPTVT--TTEYWSQ------- 662
Db       516 DISTESLMTAMTSTTRLTSAITSKTTLTSLKTTASRPTANSTLSSLTSSILSSTLVPSTD 575

Qy       663 ---SYATTTTITAPPGET--DTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVL 717
Db       576 MITSHTTNLTRSSPLLATLPTTITRSTPTSETTYPTSPTSTVKGSTTSIRYSTSMTGTLS 635

Qy       718 IRE--PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATT 775
Db       636 METSLPPT---------SSSLPTTETATMTP---TTTLITTTPN----TTSHSTPSFTSS 679

Qy       776 TTVTAPPGGTDTVIIYESMSSSKISTSSNDITS---IIPSFSRPHYVNS--TTSDLSTFE 830
Db       680 T------------IYSTVSTSTTAITSHFTTSETAVTPTPVTPSSLSTDIPTTSLRTLT 726

Qy       831 SSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSI 890
Db       727 PSSVGTSTSLTTTTDFPSIPTDISTLPTRTHII----------SSSPSI---QSTETSSL 773

Qy       891 VTSTVPTASTMSDSLSSTDG----------------------ISATSSDNVSKSGVSVT 927
Db       774 VGTTSPTMSTVRMTLRITENTPISSFSTSIVVIPETPTQTPPVLTSATGTQTSPAPTTVT 833

Qy       928 ---TETSVTTIQT-TPNPLSSSVTSLTQ----------LSSIPSVSESESKVTFTSNGDN 973
Db       834 FGSTDSSTSTLHTLTPSTALSTIVSTSQVPIPSTHSSTLQTTPSTPSLQTSLTSTSEFTT 893

Qy       974 QSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTD---LTSEPTNTREQPTTLSTTSNSI- 1029
Db       894 ESFTRGSTSTNA---ILTSFST-----IIWSSTPTIIMSSSPSSASITP-VFSTTIHSVP 944

Qy      1030 -------TEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLK- 1081
Db       945 SSPYIFSTENVGSASITG-----------FPSLSSSATTSTS---STSSSLTTALTEITP 990

Qy      1082 ------PSMGENSGLTTSTEIEATTTSP-------TEAPSPAVSS------GTDVTTEPT 1122
Db       991 FSYISLPSTTPCPGTITITIVPASPTDPCVEMDPSTEATSPPTTPLTVFPFTTEMVTCPT 1050

Qy      1123 DTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDL--TSSLTTGTSASTS----ANS 1176
Db       1051 SISIQ-TTLTTYMDTSSMMPESESSISPNASSSTGTGTVPTNTVFTSTRLPTSETWLSNS 1109

Qy      1177 ELV----------------TSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQ 1220
Db       1110 SVIPLPLPGVSTIPLTMKPSSSLPTILRTSSKSTHPSPPTTRTSETPVATTQTPTTLTSR 1169

Qy      1221 VTS------SSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLF 1259
Db       1170 RTTRITSQMTTQSTLTTTAGTCDNGG------TWEQGQCACLPGF 1208
```

RESULT 9
AAY95559
ID AAY95559 standard; Protein; 2870 AA.
XX
AC AAY95559;
XX
DT 10-OCT-2000 (first entry)
XX
DE Caenorhabditis elegans LOV-1 deletion allele lov-1(sy582del).
XX
KW Location of vulva; LOV-1 gene; nematode; mating behaviour;
KW polycystin; polycystic kidney disease; animal model;
KW signal transduction; mutant; mutein.
XX
OS Caenorhabditis elegans.
OS Synthetic.
XX
PN WO200040711-A2.
XX
PD 13-JUL-2000.
XX
PF 06-JAN-2000; 2000WO-US00521.
XX
PR 06-JAN-1999; 99US-0115127.
XX
PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
XX
PI Sternberg PW, Barr MM;
XX
DR WPI; 2000-452537/39.
XX
PT Isolated nucleic acid molecules from Caenorhabditis elegans useful for
PT producing transgenic nematodes with altered mating behavior for
PT identifying genes or regulatory factors involved in polycystic kidney
PT disease -
XX
PS Example 1; Page 131-139; 142pp; English.
XX
CC The present sequence is that of a deletion allele, termed
CC lov-1(sy582del), of the Caenorhabditis elegans LOV-1 protein
CC (see AAY95556), generated by genomic deletion of the PKD/channel
CC domain of the lov-1 gene by PCR. LOV-1 is an orthologue of
CC the human polycystin-1 (PKD1) gene that is defective in human
CC autosomal dominant polycystic kidney disease (PKD). The LOV-1
CC protein is involved in chemosensory or mechanosensory signal
CC transduction in sensory neurons. It is required for 2 male
CC sensory behaviours, 'response' of males to hermaphrodites, and
CC 'location' of the vulva. Nematodes such as C. elegans that
CC express mutant or wild-type LOV-1 or PKD-2 (see AAY95557) can be used
CC to study the functions of the proteins encoded by these genes, to
CC screen for other genes involved in PKD, to identify mutations
CC involved in the disease, and to screen for drugs that affect PKD.
CC Behaviours controlled by the action of the genes or gene products
CC are identified and used in the assays. Hence, an animal model is
CC provided that permits study of the etiology of PKD and provides a
CC tool to identify the genes involved in the disease pathway, to

```
CC   identify compounds that may be used to treat or alter the disease
CC   progression, lessen its severity or ameliorate symptoms.
XX
SQ   Sequence   2870 AA;

  Query Match             8.5%;  Score 553.5;  DB 21;  Length 2870;
  Best Local Similarity   24.3%;  Pred. No. 5.8e-20;
  Matches  290;  Conservative  154;  Mismatches  401;  Indels  349;  Gaps   51;

Qy       252 YTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLR 311
Db       194 YREKCESGEINEEYARRMC-KRPYRSE--KSTAISDSQGVYYDGQVLKGVR--AKQFSMR 248

Qy       312 WTG-----YKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTT 366
Db       249 TSGSPTLRRMKRDAGDN------TCDYTIESTSTSTTTPTTTTVTSTVT-------STTT 295

Qy       367 ITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 426
Db       296 VPTSTSTVTTAMSTSTSTPSTSTTI-----ESTSTTFTS----TASTSTSSTSTTQQSSS 346

Qy       427 VVVQVPLPNPTVSTTEYWSQSFATTTT--VTAPPGGTDTVIIREPPNHTVTTTEYWSQSF 484
Db       347 TITSSP-SSTTLST------SIPTTTTPEITSTLSSLPDNAICSYLDETTTSTTFTTTML 399

Qy       485 ATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNP 544
Db       400 TSTTT--------------EEPSTSTTTTEVTSTSSTVTTT--------------EPTTT 431

Qy       545 TVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGG 604
Db       432 LTTSTASTSTTEPSTSTVTTSPS-----TSPVTSTVTS------SSSSSTTVTTP--- 475

Qy       605 TDTVIIREPPNHTVTTTEYWSQSFATTTTVTGP------PSGTDTVIIREPPNPTVTTTE 658
Db       476 TSTESTSTSPSSTVTTS-----TTAPSTSTTGPSSSSSTPSSTASSSVSSTASSTQSSTS 530

Qy       659 YWSQSYATTT---TITAPPGETDTVLIREPPNHTVTTTEYWSQS-------------YAT 702
Db       531 --TQQSSTTTKSETTTSSDGTNPDFYFVE----KATTTFYDSTSVNLTLNSGLGIIGYQT 584

Qy       703 TTTVTAPPGETDTVLIRE------------------------PPNHTVTTT--------- 729
Db       585 SIECTSPTSSNYVSTTKDGACFTKSVSMPRLGGTYPASTFVGPGNYTFRATMTTDDKKVY 644

Qy       730 ----EYWSQSYATTT-----TVTAPPGGTDTVIIREPPNP-----TVTTTEYWSQSFATT 775
Db       645 YTYANVYIQEYSSTTIESESSTSAVASSTSST----PSTPSSTLSTSTVTEPSSTRSSDS 700

Qy       776 TTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMN 835
Db       701 TTTSA--GSTTT--LQESTTTSEESTTDSSTTTISDTSTSTSSPSSTTADSTS--TLSVD 754

Qy       836 TPTSISSDGMLLSST----------TLVTESETTTELICSDGKECSRLSS---------- 875
Db       755 QFDFILDSGLSWNETRHNEDSINIVPLPTNAITPTER--SQTFECRNVSTEPFLIIKEST 812

Qy       876 ------------SSGIVTNP---------------------------------DSN 886
Db       813 CLNYSNTVLNATYSSNIPIQPIETFLVGIGTYEFRINMTDLTTMQVVSHIFTLNVVADST 872

Qy       887 ESSIVTSTVPT-ASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSS 945
Db       873 STSEVTSTTSTGSSSESSAISTTSGIESTSTLEASTTDASQDSSTSTSDSGTT-----SD 927

Qy       946 VTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTST------------EIEIVTT 992
Db       928 STTIDSSNSTPSTSDSSG---LSQTPSDSSSASDSMRTTTVDPDASTETPYDFVLENLTW 984

Qy       993 SSTK-------VLPPVVSS-----NTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTG 1040
Db       985 NETVYYSENPFYITPIPNKEPGALTTAMTCQCRNDSSQPFVLLKESNCLTE-------FG 1037

Qy       1041 DNGDNTSST--NPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGEN-SGLTTSTEIE 1097
Db       1038 KNGAYSASVSFNPMTSFVPAT----------GTYEFLINVTNRASGESASHIFTMNVVL 1086

Qy       1098 ATTTSPTEAPSPAVSS---------GTDVT--TEPTDTREQPTTLSTTSKTNSELVATTQ 1146
Db       1087 PTTT--TETPPTTVSSSDDAGGKTGGTGATGGTGGTGSGGSATTLST-----GDAVRSTT 1139

Qy       1147 ATNENGGKSPSTDLTSSLTT----GTSASTSANSELVTSGSVT---------------- 1185
Db       1140 SGSGSGQSSTGSGAGGSGTTASGSGSGGSSSGTGSDGVNSGKTTALNGDGTGSGTATTPGS 1199

Qy       1186 ----GGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQV-----TSSSPSTNT 1230
Db       1200 HLGDGGSTSGSGSDSNGSSGVSTKSSSGSDTSGSSDSSGANGAFSATAQPSTRT 1253


RESULT 10
AAY95556
ID   AAY95556 standard; Protein; 3178 AA.
XX
AC   AAY95556;
XX
DT   10-OCT-2000  (first entry)
XX
DE   Caenorhabditis elegans LOV-1 (location of vulva) protein.
XX
KW   Location of vulva; LOV-1 gene; nematode; mating behaviour;
KW   polycystin; polycystic kidney disease; animal model;
KW   signal transduction.
XX
OS   Caenorhabditis elegans.
XX
PN   WO200040711-A2.
XX
PD   13-JUL-2000.
XX
PF   06-JAN-2000; 2000WO-US00521.
XX
PR   06-JAN-1999;   99US-0115127.
XX
PA   (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
PI   Sternberg PW,  Barr MM;
XX
DR   WPI; 2000-452537/39.
DR   N-PSDB; AAA50043.
XX
PT   Isolated nucleic acid molecules from Caenorhabditis elegans useful for
PT   producing transgenic nematodes with altered mating behavior for
PT   identifying genes or regulatory factors involved in polycystic kidney
PT   disease -
XX
PS   Claim 5; Page 114-123; 142pp; English.
XX
CC   The present sequence is that of the polypeptide encoded by the
CC   Caenorhabditis elegans LOV-1 gene (see AAA50043), an orthologue of
CC   the human polycystin-1 (PKD1) gene that is defective in human
CC   autosomal dominant polycystic kidney disease (PKD).  The LOV-1
CC   protein is involved in chemosensory or mechanosensory signal
CC   transduction in sensory neurons.  It is required for 2 male
CC   sensory behaviours, 'response' of males to hermaphrodites, and
CC   'location' of the vulva.   Nematodes such as C. elegans that
CC   express mutant or wild-type LOV-1 or PKD-2 (see AAY95557) can be used
CC   to study the functions of the proteins encoded by these genes, to
CC   screen for other genes involved in PKD, to identify mutations
CC   involved in the disease, and to screen for drugs that affect PKD.
CC   Behaviours controlled by the action of the genes or gene products
CC   are identified and used in the assays.  Hence, an animal model is
CC   provided that permits study of the etiology of PKD and provides a
CC   tool to identify the genes in the disease pathway, to
CC   identify compounds that may be used to treat or alter the disease
CC   progression, lessen its severity or ameliorate symptoms.
```

```
XX
SQ   Sequence   3178 AA;

  Query Match          8.5%;  Score 553.5;  DB 21;  Length 3178;
  Best Local Similarity   24.3%;  Pred. No. 6.6e-20;
  Matches  290;  Conservative  154;  Mismatches  401;  Indels  349;  Gaps   51;

Qy        252 YTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLR 311
              | : | |  |  :|       ||:      :| ::     | :     | |  :| |::|
Db        194 YREKCESGEINEEYARRMC-KRPYRSE--KSTAISDSQGVYYDGQVLKGVR--AKQFSMR 248

Qy        312 WTG-----YKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTT 366
              :|        ||| |      |    :||:  || |   :|  | |          |||
Db        249 TSGSPTLRRMKRDAGDN------TCDYTIESTSTSTTTPTTTTVTSTVT-------STTT 295

Qy        367 ITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 426
              : ||   |||:  | |:    : |:       |:|| ||    | :|:|: |: | |   :
Db        296 VPTSTSTVTTAMSTSTSTPSTSTTI-----ESTSTTFTS----TASTSTSSTSTTQQSSS 346

Qy        427 VVVQVPLPNPTVSTTEYWSQSFATTTT--VTAPPGGTDTVIIREPPNHTVTTTEYWSQSF 484
              :   |  : |:||      |  |||| :|:         |    : | |:| : :
Db        347 TITSSP-SSTTLST------SIPTTTTPEITSTLSSLPDNAICSYLDETTTSTTFTTTML 399

Qy        485 ATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNP 544
              :|||                | |: : ||||  | |   |||             ||
Db        400 TSTTT-------------EEPSTSTTTTEVTSTSSTVTTT-------------EPTTT 431

Qy        545 TVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGG 604
              |:|   | :  :|:|||  |          |   |||:      | :::|||| |
Db        432 LTTSTASTSTTEPSTSTVTTSPS-------TSPVTSTVTS------SSSSSTTVTTP--- 475

Qy        605 TDTVIIREPPNHTVTTTEYWSQSFATTTTVTGP------PSGTDTVIIREPPNPTVTTTE 658
              | |      |: ||||:     : | :|: |||      || | :  :    : | ::|
Db        476 TSTESTSTSPSSTVTTS-----TTAPSTSTTGPSSSSSTPSSTASSSVSSTASSTQSSTS 530

Qy        659 YWSQSYATTT---TITAPPGETDTVLIREPPNHTVTTTEYWSQS-------------YAT 702
              :|  :|||   | |:  |       |      ||| | | |             | |
Db        531 --TQQSSTTTKSETTTSSDGTNPDFYFVE----KATTTFYDSTSVNLTLNSGLGIIGYQT 584

Qy        703 TTTVTAPPGETDTVLIRE-----------------------PPNHTVTTT--------- 729
              :   |:|         ::                      | |:|   |
Db        585 SIECTSPTSSNYVSTTKDGACFTKSVSMPRLGGTYPASTFVGPGNYTFRATMTTDDKKVY 644

Qy        730 ----EYWSQSYATTT-----TVTAPPGGTDTVIIREPPNP-----TVTTTEYWSQSFATT 775
                  : | |::||     : :|   | :     |  |     | | ||  |   : :
Db        645 YTYANVYIQEYSSTTIESESSTSAVASSTSST----PSTPSSTLSTSTVTEPSSTRSSDS 700

Qy        776 TTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMN 835
              || :|  | | |  : || ::|: ||: :  |:|  : :     :|||:| ::  : |::
Db        701 TTTSA--GSTTT--LQESTTTSEESTTDSSTTTISDTSTSTSSPSSTTADSTS--TLSVD 754

Qy        836 TPTSISSDGMLLSST----------TLVTESETTTELICSDGKECSRLSS---------- 875
                  |   |:  : |          | | : | ||   |   ||  :|:
Db        755 QFDFILDSGLSWNETRHNEDSINIVPLPTNAITPTER--SQTFECRNVSTEPFLIIKEST 812

Qy        876 ------------SSGIVTNP-----------------------------------DSN 886
                         || |   |                                    ||
Db        813 CLNYSNTVLNATYSSNIPIQPIETFLVGIGTYEFRINMTDLTTMQVVSHIFTLNVVADST 872

Qy        887 ESSIVTSTVPT-ASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSS 945
              :| ||||  | :|: | ::|:| || :||:   | :  |  : || :   ||     |
Db        873 STSEVTSTTSTGSSSESSAISTTSGIESTSTLEASTTDASQDSSTSTSDSGTT-----SD 927

Qy        946 VTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTST------------EIEIVTT 992
              |::   :| || |:|      :   : |   ||  |:|             :| :|   .
Db        928 STTIDSSNSTPSTSDSSG---LSQTPSDSSSASDSMRTTTVDPDASTETPYDFVLENLTW 984

Qy        993 SSTK-------VLPPVVSS-----NTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTG 1040
              : |         : |: :       | :| :  |   ||  |   || :||        |
Db        985 NETVYYSENPFYITPIPNKEPGALTTAMTCQCRNDSSQPFVLLKESNCLTE-------FG 1037

Qy       1041 DNGDNTSST--NPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGEN-SGLTTSTEIE 1097
               ||  ::|   ||: :   :|           |::|       : ||: | : |   :

Db       1038 KNGAYSASVSFNPMTSFVPAT----------GTYEFLINVTNRASGESASHIFTMNVVL 1086

Qy       1098 ATTTSPTEAPSPAVSS---------GTDVT--TEPTDTREQPTTLSTTSKTNSELVATTQ 1146
                 |||  || |   |||          ||  |  |  | :    |||||      : | :|
Db       1087 PTTT--TETPPTTVSSSDDAGGKTGGTGATGGTGGTGSGGSATTLST-----GDAVRSTT 1139

Qy       1147 ATNENGGKSPSTDLTSSLTT----GTSASTSANSELVTSGSVT---------------- 1185
                 : : :|  |  :    | ||    |:  |:   |: | ||  |
Db       1140 SGSGSGQSSTGSGAGGSGTTASGSGSGGSSGTGSDGVNSGKTTALNGDGTGSGTATTPGS 1199

Qy       1186 ----GGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQV-----TSSSPSTNT 1230
                   ||: : : :| : |: |:  :|  |:|  :: |         :: ||| |
Db       1200 HLGDGGSTSGSGSDSNGSSGVSTKSSSGSDTSGSSDSSGANGAFSATAQPSTRT 1253


RESULT 11
AAY59288
ID   AAY59288 standard; Protein; 957 AA.
XX
AC   AAY59288;
XX
DT   25-APR-2000  (first entry)
XX
DE   Human MUC11 polypeptide.
XX
KW   Mucin; MUC11; MUC12; human; chromosome 7q22; epithelial inflammation;
KW   Crohn's disease; ulcerative colitis; asthma; chronic bronchitis;
KW   colorectal cancer; cystic fibrosis; inflammatory bowel disease;
KW   breast cancer.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   68
FT                   /note= "potential N-glycosylation site"
XX
PN   WO200004142-A1.
XX
PD   27-JAN-2000.
XX
PF   16-JUL-1999;   99WO-AU00579.
XX
PR   16-JUL-1998;   98AU-0004708.
XX
PA   (COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.
PA   (ORDE-) ORDER OF SISTERS OF MERCY IN QUEENSLAND.
XX
PI   Williams SJ,  Antalis TM,  Mcguckin MA,  Gotley DC;
XX
DR   WPI; 2000-182416/16.
DR   N-PSDB; AAZ58824.
XX
PT   Novel MUC nucleic acid corresponding to mucin gene, useful for treating
PT   associated disease conditions e.g. colorectal, breast cancer, cystic
PT   fibrosis and inflammatory bowel disease  -
XX
PS   Claim 10; Page 83-88; 103pp; English.
XX
CC   The invention provides mucin genes (MUC11 and MUC12) located on human
CC   chromosome 7q22. The mucin genes or its portion is used in detecting
CC   polymorphism, mutation, deletion, truncation and expansion in the gene
CC   or its gene transcript. Pharmaceutical compositions and gene therapy
CC   constructs comprising the mucin genes are used for treating disease
CC   conditions associated with aberrant Mucin expression, altered properties
CC   of mucus or epithelial inflammatory processes involving Mucins like
CC   Crohn's disease, ulcerative colitis, asthma, chronic bronchitis and
CC   colorectal cancer, cystic fibrosis, inflammatory bowel disease and breast
CC   cancer. The mucin genes and the polypeptides are used for determining
CC   these diseases or their predisposition. The MUC11 and MUC12 polypeptides
CC   are used for preparing antagonist and antibodies. The present sequence
CC   represents the human MUC11 polypeptide.
XX
SQ   Sequence   957 AA;
```

```
  Query Match             8.4%;  Score 545.5;  DB 21;  Length 957;
  Best Local Similarity   25.9%;  Pred. No. 3.8e-20;
  Matches  266;  Conservative  139;  Mismatches  406;  Indels  215;  Gaps   43;

Qy        338 TTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPY 397
              |||       | | :  |       | :| ||   | ||:     : :|  :|     |
Db          6 TTAFPGSTTMPGVSQESTAS--HSSPGSTDTTLSPGSTTA-----SSLGPESTTFHSGPG 58

Qy        398 HTTTTVTSEWT---GTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTT--------EYWSQ 446
              | ||:  : |   | :  :|   : | |  | :       :|  |||         : |
Db         59 STETTLLPDNTTASGLLEASTPVHSSTGSPHTTL------SPAGSTTRQGESTTFQSWPN 112

Qy        447 SFATTTTVTAPPGGTDTVIIREPPNH---TVTTTEYWSQSFAT------TTTVTAPPGGT 497
              |  ||    |||  |   :     :|   : | | ::| |  |      :||| : |  |
Db        113 SKDTT---PAPPTTTSAFVELSTTSHGSPSSTPTTHFSASSTTLGRSEESTTVHSSPVAT 169

Qy        498 DSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYA 557
              :      |:|  :||   |     :||  : || | ::   |    :    |  :  :
Db        170 ATT-----PSPARSTT---SGLVEESTTYHSSPGSTQTMHFPESDTTSGRGEESTTSHSS 221

Qy        558 TTTTVTAPPGGTDSVIIREP------PNHTVTT----------TEYWSQSYA---TTTT 597
              || |::: |  | | :: ||      |  | ||          :|  : |::    | |
Db        222 TTHTISSAP-STTSALVEEPTSYHSSPGSTATTHFPDSSTTSGRSEESTASHSNQDATGT 280

Qy        598 VTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATT----TTVTGPPSGTDTVIIREPPNPT 653
              :  |   | :|:: |      ||:   | |  ||    :| |   |   |     | :|
Db        281 IVLPARSTTSVLLGES-----TTSPISSGSMETTALPGSTTTPGLSEKSTTFHSSPRSPA 335

Qy        654 VTTTEYWSQSYAT---TTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPP 710
              | :   : |      :||  : || | |    :      |||   |:    :|| : |
Db        336 TTLSPASTTSSGVSEESTTSHSRPGSTHTTAFPDS-----TTTPGLSRH---STTSHSSP 387

Qy        711 GETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQ 770
              | ||| |:  | :   |||   ||    :||  : || ||| :     :|  ||   : |
Db        388 GSTDTTLL--PAS---TTTSGPSQE---STTSHSSPGSTDTAL-----SPGSTTALSFGQ 434

Qy        771 SFATTTTVTAPPGGTDTVIIYESMSSSKI--------STSSNDITSIIPSFSRPHYVNST 822
                :||  : || | | :  :| :|| |         |:: :  |:: |: |    :
Db        435 E--STTFHSSPGSTHTTLFPDSTTSSGIVEASTRVHSSTGSPRTTLSPASSTSPGLQGE 491

Qy        823 TSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTN 882
              ::   |  :|:  ||::         ||      |:||          :   :||   |
Db        492 STAFQTHPASTHTTPST--------PSTATAPVEESTTYHRSPSSTPTTHFPASS---TT 540

Qy        883 PDSNESSIVTSTVPTAS--TMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPN 940
              :| | :  : | ||  | | : |:| |   :::  :|     :||    |   |||
Db        541 SGHSEKSTIFHSSPDASGTTPSSAHSTTSGRGESTTSRISPGSTEITTLPGST---TTPG 597

Qy        941 PLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPP 1000
              || : |:       |: : | : :|    |: :| |  ||  ||   :   |:|    |
Db        598 -LSEASTTFYSSPRSPTTTLSPASMTSLGVGE-ESTTSRSQPGSTHSTVSPASTT---TP 652

Qy       1001 VVS--SNTDLTSEPTNTR-------------EQPTTL-----STTSNSITEDITTS--- 1036
              :|  | |  :| | :|               |:|||      ||  :   ||| |||
Db        653 GLSEESTTVYSSSPGSTETTVFPRSTTTSVRGEEPTTFHSRPASTHTTLFTEDSTTSGLT 712

Qy       1037 -QPTGDNGDNTSSTNPVPTVATSTLASASEEDNK-SGSHESASTSLKPSMGENSGL---- 1090
              : |   |   |:   :|  || | |   ||      |  |  |:| |:    |||
Db        713 EESTAFPGSPASTQTGLP--ATLTTADLGEESTTFPSSSGSTGTTLSPARSTTSGLVGES 770

Qy       1091 ------------------TTSTEIEATTT------SPTEAPSPAVSSGTDVTTEPTDTR 1125
                               || :  | :||      ||    ||| :: : |: | : :
Db        771 TPSRLSPSSTETTTLPGSPTTPSLSEKSTTFYTSPRSPDATLSPATTTSSGVSEESSTSH 830

Qy       1126 EQPTTLSTTS--------------KTNSELVATTQATNENGGKSPSTDLTSSL------- 1164
              || :  ||:              ||:     :|:||      || :   |||
Db        831 SQPGSTHTTAFPDSTTTSGLSQEPKTSHSSQGSTEAT-----LSPGSTTASSLGQQSTTF 885

Qy       1165 --TTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVT 1222
                : | : :|    : :||| |       :|    |:|    |::      ::|  |
Db        886 HSSPGDTETTLLPDDTITSGLVEASTPTHSSTGSLHTTLTPASSTSAGLQEESTTFQSWP 945

Qy       1223 SSSPST 1228
                  ||| :|
Db        946 SSSDTT 951
```

RESULT 12
AAM24513
ID AAM24513 standard; Protein; 957 AA.
XX
AC AAM24513;
XX
DT 12-OCT-2001 (first entry)
XX
DE C900P predicted amino acid sequence.
XX
KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW immunogenic; gene therapy; vaccine; colonic cancer.
XX
OS Homo sapiens.
XX
PN WO200149716-A2.
XX
PD 12-JUL-2001.
XX
PF 29-DEC-2000; 2000WO-US35596.
XX
PR 30-DEC-1999; 99US-0476296.
PR 10-JAN-2000; 2000US-0480321.
PR 15-FEB-2000; 2000US-0504629.
PR 06-MAR-2000; 2000US-0519444.
PR 19-MAY-2000; 2000US-0575251.
PR 29-JUN-2000; 2000US-0609448.
PR 28-AUG-2000; 2000US-0649811.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
DR WPI; 2001-441847/47.
XX
PT Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer -
XX
PS Claim 2; Page 437-440; 472pp; English.
XX
CC The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be
CC used to treat disorders associated with decreased expression by
CC rectifying mutations or deletions in a patient's genome that affect the
CC activity of TCAPs by expressing inactive proteins or to supplement the
CC patients own production of them. Additionally, (II) may be used to
CC produce the TCAP proteins, by inserting the nucleic acids into a host
CC cell culturing the cell to express the protein. (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC polymerase chain reaction (PCR) and hybridisation assays to detect and
CC quantitate the presence of similar nucleic acids in samples, and
CC therefore which patients may be in need of restorative therapy. (I) may
CC also be used as antigens in the production of antibodies against TCAPs
CC and in assays to identify modulators of TCAP expression and activity.
CC Anti-(I) antibodies and antagonists may also be used to down regulate
CC TCAP expression and activity. The anti-(I) antibodies may also be used
CC as diagnostic agents for detecting the presence of TCAPs in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC given in the exemplification of the present invention.
XX
SQ Sequence 957 AA;

```
  Query Match             8.4%;  Score 545.5;  DB 22;  Length 957;
  Best Local Similarity   25.9%;  Pred. No. 3.8e-20;
  Matches  266;  Conservative  139;  Mismatches  406;  Indels  215;  Gaps   43;

Qy        338 TTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPY 397
              |||       | | :  |       | :| ||   | ||:     : :|  :|     |
Db          6 TTAFPGSTTMPGVSQESTAS--HSSPGSTDTTLSPGSTTA-----SSLGPESTTFHSGPG 58

Qy        398 HTTTTVTSEWT---GTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTT--------EYWSQ 446
              | ||:  : |   | :  :|   : | |  | :        :|  |||        : |
Db         59 STETTLLPDNTTASGLLEASTPVHSSTGSPHTTL------SPAGSTTRQGESTTFQSWPN 112

Qy        447 SFATTTTVTAPPGGTDTVIIREPPNH---TVTTTEYWSQSFAT------TTTVTAPPGGT 497
              |  ||    |||  |   :     :|   : | | ::| |  |      :||| : |  |
Db        113 SKDTT---PAPPTTTSAFVELSTTSHGSPSSTPTTHFSASSTTLGRSEESTTVHSSPVAT 169

Qy        498 DSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYA 557
              :      |:|  :||   |     :||  : || | ::   |    :    |  :   :
Db        170 ATT-----PSPARSTT---SGLVEESTTYHSSPGSTQTMHFPESDTTSGRGEESTTSHSS 221

Qy        558 TTTTVTAPPGGTDSVIIREP------PNHTVTT-----------TEYWSQSYA---TTTT 597
              || |::: |  | | :: ||      |  | ||           :|  : |::    | |
Db        222 TTHTISSAP-STTSALVEEPTSYHSSPGSTATTHFPDSSTTSGRSEESTASHSNQDATGT 280

Qy        598 VTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATT----TTVTGPPSGTDTVIIREPPNPT 653
              :  |   | :|:: |      ||:   | |  ||    :| |   |    |     | :|
Db        281 IVLPARSTTSVLLGES-----TTSPISSGSMETTALPGSTTTPGLSEKSTTFHSSPRSPA 335

Qy        654 VTTTEYWSQSYAT--TTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPP 710
              | :   : |      :||  : || | |    :      |||   |:    :||  : |
Db        336 TTLSPASTTSSGVSEESTTSHSRPGSTHTTAFPDS-----TTTPGLSRH---STTSHSSP 387

Qy        711 GETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQ 770
              | ||| |:  | :   |||  ||    :||  : || ||| :     :|  ||   : |
Db        388 GSTDTTLL--PAS---TTTSGPSQE---STTSHSSPGSTDTAL-----SPGSTTALSFGQ 434

Qy        771 SFATTTTVTAPPGGTDTVIIYESMSSSKI--------STSSNDITSIIPSFSRPHYVNST 822
                :||  : || | |  :  :| :|| |        |:: :  |:: |: |    :
Db        435 E---STTFHSSPGSTHTTLFPDSTTSSGIVEASTRVHSSTGSPRTTLSPASSTSPGLQGE 491

Qy        823 TSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTN 882
              ::   |  :|:  ||::          ||      |:||           :   :||   |
Db        492 STAFQTHPASTHTTPST--------PSTATAPVEESTTYHRSPSSTPTTHFPASS---TT 540

Qy        883 PDSNESSIVTSTVPTAS--TMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPN 940
                :| | :  : | ||  | | : |:| |   :::  :|     :||    |  |||
Db        541 SGHSEKSTIFHSSPDASGTTPSSAHSTTSGRGESTTSRISPGSTEITTLPGST---TTPG 597

Qy        941 PLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPP 1000
              || : |:       |: : | : :|    |: :| |  ||  ||   :   |:|    |
Db        598 -LSEASTTFYSSPRSPTTTLSPASMTSLGVGE-ESTTSRSQPGSTHSTVSPASTT---TP 652

Qy       1001 VVS--SNTDLTSEPTNTR-------------EQPTTL-----STTSNSITEDITTS--- 1036
              :|  | |  :| | :|               |:|||      || :   |:| |||
Db        653 GLSEESTTVYSSSPGSTETTVFPRSTTTSVRGEEPTTFHSRPASTHTTLFTEDSTTSGLT 712

Qy       1037 -QPTGDNGDNTSSTNPVPTVATSTLASASEEDNK-SGSHESASTSLKPSMGENSGL---- 1090
              : |   |   |:   :|  || | |   ||       |  |  |:| |:    |||
Db        713 EESTAFPGSPASTQTGLP--ATLTTADLGEESTTFPSSSGSTGTTLSPARSTTSGLVGES 770

Qy       1091 ------------------TTSTEIEATTT------SPTEAPSPAVSSGTDVTTEPTDTR 1125
                               || :  | :||      ||    ||| :: : |: | : :
Db        771 TPSRLSPSSTETTTLPGSPTTPSLSEKSTTFYTSPRSPDATLSPATTTSSGVSEESSTSH 830

Qy       1126 EQPTTLSTTS-------------KTNSELVATTQATNENGGKSPSTDLTSSL------- 1164
              || :  ||:             ||:     :|:||      || :   |||
Db        831 SQPGSTHTTAFPDSTTTSGLSQEPKTSHSSQGSTEAT-----LSPGSTTASSLGQQSTTF 885

Qy       1165 --TTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVT 1222
                : | : :|    : :||| |       :|     |:|    |::      ::|  |
Db        886 HSSPGDTETTLLPDDTITSGLVEASTPTHSSTGSLHTTLTPASSTSAGLQEESTTFQSWP 945

Qy       1223 SSSPST 1228

              ||| :|
Db        946 SSSDTT 951


RESULT 13
AAW06725
ID   AAW06725 standard; Protein; 849 AA.
XX
AC   AAW06725;
XX
DT   07-FEB-1997  (first entry)
XX
DE   FLO1 protein, involved in flocculation, derived from S.cerevisiae.
XX
KW   Flocculation; aggregation; floating; yeast; beer; Saccharomyces;
KW   fermentation; ds.
XX
OS   Saccharomyces cerevisiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2550
FT                   /*tag= a
FT                   /note= "partial sequence only, no start codon"
XX
PN   JP08205900-A.
XX
PD   13-AUG-1996.
XX
PF   01-FEB-1995;   95JP-0015453.
XX
PR   01-FEB-1995;   95JP-0015453.
XX
PA   (KIRI ) KIRIN BREWERY KK.
XX
DR   WPI; 1996-419830/42.
DR   N-PSDB; AAT45666.
XX
PT   DNA molecule involved in yeast flocculation - useful for determining
PT   if a sample yeast has a flocculation activity
XX
PS   Claim 1; Page 6-10; 16pp; Japanese.
XX
CC   AAW06725 is the FLO1 protein derived from a partial gene sequence
CC   from Saccharomyces cerevisiae. The gene was named Lg-FLO1. The
CC   FLO1 protein is believed to be involved in yeast flocculation
CC   (aggregation). The DNA and protein are used in a method for
CC   determining whether or not a sample yeast has flocculating activity,
CC   i.e. to distinguish if a yeast is a flocculating or floating yeast.
XX
SQ   Sequence   849 AA;

  Query Match             8.2%;  Score 534;  DB 17;  Length 849;
  Best Local Similarity   26.0%;  Pred. No. 1.3e-19;
  Matches  244;  Conservative  152;  Mismatches  386;  Indels  156;  Gaps   39;

Qy        380 TKTAPIGETATVIVDVP--YHTTTTVTSEWTGTIT------TTTTRTN--PTDSIDTVVV 429
              |   |  ||  ::|  |   :|  | |  |||| |      || | ||  |||    :|:
Db          2 TNGQPTDET-VIVVKTPTTANTIVTTTKPWTGTFTSTSTEMTTVTGTNGQPTDE-TVIVI 59

Qy        430 QVPLPNPTVS-TTEYWSQSFATT----TTVTAPPG-GTD--TVIIREPPNH--TVTTTEY 479
              : |     :| ||| |: :| :|    ||:|   |  ||   ::|| | :     ||||
Db         60 RTPTSEGLISTTTEPWTGTFTSTSTEVTTITGTNGQPTDETVIVIRTPTSEGLVTTTTEP 119

Qy        480 WSQSFATT----TTVTAPPG--GTDSVIIREPPNPTVTTTEYWSQSFATTTTVTA----- 528
              |: :| :|    ||:|   |    ::||: : |   ::::   | |   |:::|:
Db        120 WTGTFTSTSTGMTTITGTNGLPTDETVIVVKTPTTAISSSLSLSSSGQITSSITSSRPII 179

Qy        529 ----PPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTT 584
                  |  ||   :|      :  |:   : |  |:: ||:      |||     :  ||:
Db        180 TPFYPSNGTS--VISSSDTSSSDTSSLVTSSLVTSSLVTS------SVISSSVTSSLVTS 231

Qy        585 TEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTV 644
              :     |  |:: ||:      :||     :   |::   | |  :::|       | |
```

```
Db        232 SVI--SSSVTSSLVTS------SVISSSVTSSLFTSSPVISSSVISSSTT------TSTS 277

Qy        645 IIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTT 704
Db        278 IFFESSKSSVIPTSSSTSGSSESKTSSASSSSSSSSISSESPKSTYSSSSLPPVTSATTS 337

Qy        705 TVTAP--PGETDTVLIRE-----PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREP 757
Db        338 QETASSLPPVTSVTTSQEITSSLPPVISTTTSQETASSLPPVTSATTSQETASSL----P 393

Qy        758 PNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPH 817
Db        394 PVTSVTTSQEITSSLPPVTSVTT------SQEITSSLPPVTSATTSQEITSSLPP----- 442

Qy        818 YVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSS 877
Db        443 -VTSATTSQET--ASSLPPVTSAATSQETASSLPPVTSVTTSQE-----------ITSSL 488

Qy        878 GIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGI-----SATSSDNVSKSGVSVTTETSV 932
Db        489 PPVTSVTTSQE--ITSSLPPVTSVTTSQEITSSLPPVTSATTSQETASSLPPATTTKTSE 546

Qy        933 TTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTT 992
Db        547 QTTLVTVTSCESHVCTESISSAIVSTA------TVTVSGVTTEYTTWCPISTTEITKQTT 600

Qy        993 SSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPV 1052
Db        601 ETTKQTKGTTKQTKGTTEQTTETTKQ-TTVVTISSCESDICSKTASPAIVSTSTATINDV 659

Qy       1053 PTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAP---SP 1109
Db        660 TTEYT-TWCPISTTESK---------------QQTTLVTVTSCESGVCSETASPAIVST 702

Qy       1110 AVSSGTDVTT-----EPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSL 1164
Db        703 ATATVNDVVTVYPTWRPQTTNEQ----SVSSKMNS---ATSETTTNTGAAETKTAVTSSL 755

Qy       1165 TTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSS 1224
Db        756 SRFNHAETQTASATDVIGH--SSSVVSVSETGNTKGLITSELSTMSQQPRSTPASTMVGS 813

Qy       1225 SPSTNTFIASTYDGSGSIIQHSTWLYGLITL--LSLFI 1260
Db        814 --STASLEISTYVGIAN---------GLLTNNGISVFI 840
```

```
RESULT 14
AAR47578
ID   AAR47578 standard; Protein; 894 AA.
XX
AC   AAR47578;
XX
DT   19-JUL-1994  (first entry)
XX
DE   Flocculation protein of Saccharomyces cerevisiae.
XX
KW   Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;
KW   Major cell wall protein; glycosyl-phosphatidyl-inositol;
KW   anchoring protein; alpha factor; alpha-agglutinin; invertase;
KW   inulinase; alpha-amylase; Saccharomyces cerevisiae;
KW   floculation protein; enzymatic process; fermentation;
KW   biodegradation; catalysis.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO9401567-A.
XX
PD   20-JAN-1994.
XX
PF   07-JUL-1993;   93WO-EP01763.
XX
PR   08-JUL-1992;   92EP-0202080.
PR   14-DEC-1992;   92EP-0203899.
XX
PA   (UNIL ) UNILEVER NV.
PA   (UNIL ) UNILEVER PLC.
XX
PI   Klis FM,  Schreuder MP,  Toschka H,  Verrips CT;
XX
DR   WPI; 1994-035071/04.
DR   N-PSDB; AAQ54029.
XX
PT   Immobilisation of enzymes to microbial cell wall - by prodn. of
PT   fusion protein of enzyme linked to anchoring protein
XX
PS   Example 10; Page 59-64; 99pp; English.
XX
CC   The flocculation protein is used in a method to immobilise enzymes
CC   to a microbial cell wall.  The coding sequence (FLO 1) is used in
CC   the production of a recombinant polynucleotide which comprises a
CC   structural gene encoding a protein with catalytic activity and at
CC   least part of a gene encoding at least the C-terminus of a protein
CC   capable of anchoring in a eukaryotic or prokaryotic cell wall.  The
CC   anchoring fragment or protein is selected from alpha agglutinin,
CC   AGA 1, FLO 1, major cell wall protein of lower eukaryotes or a
CC   proteinase of lactic acid bacteria.  The recombinant polynucleotide
CC   preferably also comprises a sequence encoding a signal peptide to
CC   ensure secretion of the expressed product.  The signal peptide is
CC   preferably derived from glycosyl-phosphatidyl-inositol, anchoring
CC   protein, alpha factor, alpha-agglutinin, invertase or inulinase,
CC   alpha-amylase of Bacillus or proteinases of lactic acid bacteria.
CC   The host microorganism can be used for performing enzymatic
CC   processes on an industrial scale.
XX
SQ   Sequence   894 AA;

  Query Match             7.9%;  Score 516;  DB 15;  Length 894;
  Best Local Similarity   26.3%;  Pred. No. 1.2e-18;
  Matches  251;  Conservative  126;  Mismatches  310;  Indels  266;  Gaps   43;
```

```
Qy        437 TVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTT------------------ 477
Db         53 TYSNAAYMAYGYASKTKLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGA 112

Qy        478 -------EYWSQS-FATTTTVTAPPGGTDSVIIREPPNPTVTTTEY--------WSQSFA 521
Db        113 CSNSQGIAYWSTDLFGFYTT--------------PTNVTLEMTGYFLPPQTGSYTFKFA 157

Qy        522 TT-TTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNH 580
Db        158 TVDDSAILSVGGATAFNCCAQQQPPITSTNF-------TIDGIKPWGGS------LPPNI 204

Qy        581 TVTTTEYWSQSY-------------ATTTTVTAPPGGT------------------DTVI 609
Db        205 EGTVYMYAGYYYPMKVVYSNAVSWGTLPISVTLPDGTTVSDDFEGYVYSFDDDLSQSNCT 264

Qy        610 IREPPNH----TVTTTEYWSQSFATT----TTVTGP---PSGTDTVIIREPPNPTV--TT 656
Db        265 VPDPSNYAVSTTTTTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTT 324

Qy        657 TEYWSQSYATT----TTITAPPGE-TD--TVLIREPPNHTV--TTTEYWSQSYATT---- 703
Db        325 TEPWTGTFTSTSTEVTTITGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEM 384

Qy        704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNP--T 761
Db        385 TTVTGTNG--------QPTDETV-------------------------IVIRTPTSEGLV 411

Qy        762 VTTTEYWSQSFATTTTVTAPPGGT------DTVIIYE----SMSSSKISTSSNDITSIIP 811
Db        412 TTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSSSGQITSSIT 471

Qy        812 SFSR----PHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDG 867
Db        472 S-SRPIITPFYPSNGTSVISSSVISSSVTSSLFTSSPVI--SSSVISSSTTTSTSIFSES 528
```

```
Qy       868 KECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVT 927
Db       529 SKSSVIPTSSSTSGSSESETSS-----------AGSVSSSSFISSESSKSPTYSSSSLP 576

Qy       928 TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEI 987
Db       577 LVTSATTSQETASSLPPATTTKT--------SEQTTLVTV--------------TSCES 613

Qy       988 EIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTS 1047
Db       614 HVCTES---ISPAIVSTATVTVSGVTTEYTTWCPISTTE-------TTKQTKGTT-EQTT 662

Qy      1048 STNPVPTVATSTLASASEED--NKSGSHESASTSLKPSMGENSGLTT-----STEIEATT 1100
Db       663 ETTKQTTVVT---ISSCESDVCSKTASPAIVSTSTATINGVTTEYTTWCPISTTESRQQT 719

Qy      1101 TSPT----------EAPSPA-VSSGTD-----VTTEPTDTREQPTTLSTTSKTNSELVAT 1144
Db       720 TLVTVTSCESGVCSETASPAIVSTATATVNDVVTVYPTWRPQTANEESVSSKMNSATGET 779

Qy      1145 TQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGG-----AVASASNDQSHS 1199
Db       780 T--TNTLAAETTTNTVAAETITNTGA---AETKTVVTSSLSRSNHAETQTASATDVIGHS 834

Qy      1200 TSV-----TNSNSIVSNTPQTTLSQQVTSSSP------STNTFIASTYDGSGS 1241
Db       835 SSVVSVSETGNTKSLTSSGLSTMSQQPRSTPASSMVGYSTASLEISTYAGSAT 887
```

RESULT 15
AAR58754
ID AAR58754 standard; Protein; 894 AA.
XX
AC AAR58754;
XX
DT 27-MAR-1995 (first entry)
XX
DE S. cerevisiae FLO1.
XX
KW Binding protein; immobilization; chimeric protein;
KW anchoring protein; Saccharomyces cerevisiae; flocculation;
KW FLO1 gene; scFv; single chain antibody; monoclonal antibody;
KW MAb; human chorionic gonadotropin; HCG.
XX
OS Saccharomyces cerevisiae.
XX
PN WO9418330-A.
XX
PD 18-AUG-1994.
XX
PF 10-FEB-1994; 94WO-EP00427.
XX
PR 10-FEB-1993; 93EP-0200350.
XX
PA (UNIL) UNILEVER NV.
PA (UNIL) UNILEVER PLC.
XX
PI De Geus P, Frenken LGJ, Klis FM, Toschka HY, Verrips CT;
XX
DR WPI; 1994-279751/34.
DR N-PSDB; AAQ67360.
XX
PT Immobilised binding proteins for specific cpds - obtd. by
PT expressing chimeric proteins comprising the binding protein and a
PT cell wall-anchoring protein in host cells
XX
PS Disclosure; Page 36-39; 78pp; English.
XX
CC Valuable compounds are isolated from complex mixtures by use of
CC immobilized ligands composed of an anchoring protein and a binding
CC protein. A suitable anchoring protein is yeast FLO1 (associated
CC with flocculation), and a gene encoding a chimeric scFv-FLO1 protein
CC that will anchor in the cell wall of a lower eukaryote, and which
CC binds HCG with high specificity, was produced.
XX
SQ Sequence 894 AA;

Query Match 7.9%; Score 516; DB 15; Length 894;
Best Local Similarity 26.3%; Pred. No. 1.2e-18;
Matches 251; Conservative 126; Mismatches 310; Indels 266; Gaps 43;

```
Qy       437 TVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTT------------------ 477
Db        53 TYSNAAYMAYGYASKTKLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGA 112

Qy       478 -------EYWSQS-FATTTTVTAPPGGTDSVIIREPPNPTVTTTEY--------WSQSFA 521
Db       113 CSNSQGIAYWSTDLFGFYTT---------------PTNVTLEMTGYFLPPQTGSYTFKFA 157

Qy       522 TT-TTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNH 580
Db       158 TVDDSAILSVGGATAFNCCAQQQPPITSTNF-------TIDGIKPWGGS------LPPNI 204

Qy       581 TVTTTEYWSQSY-------------ATTTTVTAPPGGT-----------------DTVI 609
Db       205 EGTVYMYAGYYYPMKVVYSNAVSWGTLPISVTLPDGTTVSDDFEGYVYSFDDDLSQSNCT 264

Qy       610 IREPPNH----TVTTTEYWSQSFATT----TTVTGP---PSGTDTVIIREPPNPTV--TT 656
Db       265 VPDPSNYAVSTTTTTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTT 324

Qy       657 TEYWSQSYATT----TTITAPPGE-TD--TVLIREPPNHTV--TTTEYWSQSYATT---- 703
Db       325 TEPWTGTFTSTSTEVTTITGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEM 384

Qy       704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNP--T 761
Db       385 TTVTGTNG--------QPTDETV-------------------------IVIRTPTSEGLV 411

Qy       762 VTTTEYWSQSFATTTTVTAPPGGT------DTVIIYE----SMSSSKISTSSNDITSIIP 811
Db       412 TTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSSSGQITSSIT 471

Qy       812 SFSR----PHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDG 867
Db       472 S-SRPIITPFYPSNGTSVISSSVISSSVTSSLFTSSPVI--SSSVISSSTTTSTSIFSES 528

Qy       868 KECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVT 927
Db       529 SKSSVIPTSSSTSGSSESETSS-----------AGSVSSSSFISSESSKSPTYSSSSLP 576

Qy       928 TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEI 987
Db       577 LVTSATTSQETASSLPPATTTKT--------SEQTTLVTV--------------TSCES 613

Qy       988 EIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTS 1047
Db       614 HVCTES---ISPAIVSTATVTVSGVTTEYTTWCPISTTE-------TTKQTKGTT-EQTT 662

Qy      1048 STNPVPTVATSTLASASEED--NKSGSHESASTSLKPSMGENSGLTT-----STEIEATT 1100
Db       663 ETTKQTTVVT---ISSCESDVCSKTASPAIVSTSTATINGVTTEYTTWCPISTTESRQQT 719

Qy      1101 TSPT----------EAPSPA-VSSGTD-----VTTEPTDTREQPTTLSTTSKTNSELVAT 1144
Db       720 TLVTVTSCESGVCSETASPAIVSTATATVNDVVTVYPTWRPQTANEESVSSKMNSATGET 779

Qy      1145 TQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGG-----AVASASNDQSHS 1199
Db       780 T--TNTLAAETTTNTVAAETITNTGA---AETKTVVTSSLSRSNHAETQTASATDVIGHS 834

Qy      1200 TSV-----TNSNSIVSNTPQTTLSQQVTSSSP------STNTFIASTYDGSGS 1241
Db       835 SSVVSVSETGNTKSLTSSGLSTMSQQPRSTPASSMVGYSTASLEISTYAGSAT 887
```

Search completed: June 11, 2003, 17:13:28
Job time : 58 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 17:11:31 ; Search time 19 Seconds
(without alignments)
2750.535 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQQFTLLFLYLSIASAKTI.....SIIQHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6465	99.5	1260	1	ALS1_CANAL	P46590 candida alb
2	4204.5	64.7	1119	1	ALS3_CANAL	O74623 candida alb
3	3194	49.2	1419	1	ALA1_CANAL	O13368 candida alb
4	1992.5	30.7	468	1	ALS2_CANAL	O74657 candida alb
5	1819	28.0	469	1	ALS4_CANAL	O74660 candida alb
6	940	14.5	1537	1	FLO1_YEAST	P32768 saccharomyc
7	932.5	14.4	1322	1	YAG3_YEAST	P39712 saccharomyc
8	721.5	11.1	5179	1	MUC2_HUMAN	Q02817 homo sapien
9	711	10.9	1075	1	FLO5_YEAST	P38894 saccharomyc
10	675.5	10.4	1367	1	AMYH_YEAST	P08640 saccharomyc
11	577	8.9	1161	1	DAN4_YEAST	P47179 saccharomyc
12	553.5	8.5	3178	1	YS89_CAEEL	Q09624 caenorhabdi
13	547.5	8.4	1609	1	FIG2_YEAST	P25653 saccharomyc
14	521.5	8.0	1140	1	YM96_YEAST	Q04893 saccharomyc
15	501	7.7	1169	1	YK82_YEAST	P36170 saccharomyc
16	464.5	7.2	670	1	VG50_HSVI1	Q00130 ictalurid h
17	446	6.9	1802	1	HKR1_YEAST	P41809 saccharomyc
18	443.5	6.8	1251	1	YQU3_CAEEL	Q09550 caenorhabdi
19	418.5	6.4	725	1	AGA1_YEAST	P32323 saccharomyc
20	415.5	6.4	1306	1	MSB2_YEAST	P32334 saccharomyc
21	409.5	6.3	1041	1	EGT2_YEAST	P42835 saccharomyc
22	406	6.3	2090	1	HFC1_MESAU	P51611 mesocricetu
23	404.5	6.2	2035	1	HFC1_HUMAN	P51610 homo sapien
24	392	6.0	797	1	VGLX_HSVEB	P28968 equine herp
25	374.5	5.8	881	1	YJH8_YEAST	P47033 saccharomyc
26	353	5.4	650	1	SAG1_YEAST	P20840 saccharomyc
27	351.5	5.4	5376	1	ZAN_MOUSE	O88799 mus musculu
28	349.5	5.4	860	1	CHI2_COCIM	P54197 coccidioide
29	347.5	5.4	1150	1	APMU_PIG	P12021 sus scrofa
30	338	5.2	1848	1	CBPA_CLOCL	P38058 clostridium
31	327	5.0	2358	1	YEEJ_ECOLI	P76347 escherichia
32	323	5.0	907	1	VGP3_EBV	P03200 epstein-bar
33	320	4.9	886	1	VGP3_EBVA8	Q07284 epstein-bar
34	312	4.8	2660	1	YEEJ_ECO57	Q8x8v7 escherichia
35	311	4.8	1253	1	DSPP_HUMAN	Q9nzw4 homo sapien
36	305	4.7	2090	1	N214_HUMAN	P35658 homo sapien
37	304.5	4.7	3396	1	PGCV_HUMAN	P13611 homo sapien
38	302	4.6	995	1	YIQ9_YEAST	P40442 saccharomyc
39	302	4.6	3562	1	PGCV_CHICK	Q90953 gallus gall
40	299.5	4.6	3381	1	PGCV_BOVIN	P81282 bos taurus
41	295.5	4.5	662	1	MUC1_XENLA	Q05049 xenopus lae
42	294	4.5	1849	1	IGA4_HAEIN	P45386 haemophilus
43	293.5	4.5	610	1	MUC4_HUMAN	Q99102 homo sapien
44	293	4.5	1199	1	N121_RAT	P52591 rattus norv
45	286	4.4	636	1	YNR6_YEAST	P53882 saccharomyc

ALIGNMENTS

RESULT 1
ALS1_CANAL
ID ALS1_CANAL STANDARD; PRT; 1260 AA.
AC P46590;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Agglutinin-like protein 1 precursor.
GN ALS1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11651 / B792;
RX MEDLINE=95272392; PubMed=7752895;
RA Hoyer L.L., Scherer S., Shatzman A.R., Livi G.P.;
RT "Candida albicans ALS1: domains related to a Saccharomyces cerevisiae
RT sexual agglutinin separated by a repeating motif.";
RL Mol. Microbiol. 15:39-54(1995).
CC -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC -!- SIMILARITY: TO YEAST SAG1.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; L25902; AAC41649.2; -.
KW Cell adhesion; Glycoprotein; Repeat; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 1260 AGGLUTININ-LIKE PROTEIN 1.
FT DOMAIN 433 792 10 X 36 AA TANDEM REPEATS.
FT REPEAT 433 468 1-1.
FT REPEAT 469 504 1-2.
FT REPEAT 505 540 1-3.
FT REPEAT 541 576 1-4.
FT REPEAT 577 612 1-5.
FT REPEAT 613 648 1-6.
FT REPEAT 649 684 1-7.
FT REPEAT 685 720 1-8.
FT REPEAT 721 756 1-9.
FT REPEAT 757 792 1-10.
FT DOMAIN 983 1152 2 X 26 AA APPROXIMATE REPEATS.
FT REPEAT 983 1043 2-1.
FT REPEAT 1092 1152 2-2.
FT DOMAIN 399 404 POLY-THR.
FT DOMAIN 408 418 POLY-THR.
FT DOMAIN 450 455 POLY-THR.
FT DOMAIN 486 491 POLY-THR.
FT DOMAIN 522 527 POLY-THR.

```
FT   DOMAIN      558    563       POLY-THR.
FT   DOMAIN      594    599       POLY-THR.
FT   DOMAIN      630    635       POLY-THR.
FT   DOMAIN      666    671       POLY-THR.
FT   DOMAIN      702    707       POLY-THR.
FT   DOMAIN      738    743       POLY-THR.
FT   DOMAIN      774    779       POLY-THR.
FT   DOMAIN      874    877       POLY-SER.
FT   CARBOHYD    471    471       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    579    579       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    687    687       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    723    723       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    820    820       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    886    886       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    918    918       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    973    973       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1045   1045       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1068   1068       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1260 AA;  132641 MW;  763D1063A2354C24 CRC64;

  Query Match           99.5%;  Score 6465;  DB 1;  Length 1260;
  Best Local Similarity  99.6%;  Pred. No. 1.6e-270;
  Matches 1255;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SRSQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TDSIDTVVVQVPSPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540

Qy        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600

Qy        601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660

Qy        661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720

Qy        721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780

Qy        781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840

Qy        841 SSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        841 SSDGMLLSSTTLVTESETTTESICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900

Qy        901 MSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 MSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE 960

Qy        961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT 1020

Qy       1021 TLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080

Qy       1081 KPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 KPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE 1140

Qy       1141 LVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHST 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 SVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHST 1200

Qy       1201 SVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 SVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260


RESULT 2
ALS3_CANAL
ID   ALS3_CANAL     STANDARD;      PRT;  1119 AA.
AC   O74623;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein 3 precursor.
GN   ALS3.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98309840; PubMed=9644209;
RA   Hoyer L.L., Payne T.L., Bell M., Myers A.M., Scherer S.;
RT   "Candida albicans ALS3 and insights into the nature of the ALS gene
RT   family.";
RL   Curr. Genet. 33:451-459(1998).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U87956; AAC39486.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1119       AGGLUTININ-LIKE PROTEIN 3.
```

```
FT   DOMAIN       433    792       10 X 36 AA TANDEM REPEATS.
FT   REPEAT       433    468       1-1.
FT   REPEAT       469    504       1-2.
FT   REPEAT       505    540       1-3.
FT   REPEAT       541    576       1-4.
FT   REPEAT       577    612       1-5.
FT   REPEAT       613    648       1-6.
FT   REPEAT       649    684       1-7.
FT   REPEAT       685    720       1-8.
FT   REPEAT       721    756       1-9.
FT   REPEAT       757    792       1-10.
FT   DOMAIN       399    404       POLY-THR.
FT   DOMAIN       450    455       POLY-THR.
FT   DOMAIN       557    563       POLY-THR.
FT   DOMAIN       593    597       POLY-THR.
FT   DOMAIN       630    635       POLY-THR.
FT   DOMAIN       666    671       POLY-THR.
FT   DOMAIN       702    707       POLY-THR.
FT   DOMAIN       738    743       POLY-THR.
FT   DOMAIN       774    777       POLY-THR.
FT   DOMAIN      1044   1047       POLY-THR.
FT   CARBOHYD     471    471       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     543    543       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     579    579       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     651    651       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     687    687       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     723    723       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     759    759       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     845    845       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     987    987       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1050   1050       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    1061   1061       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1119 AA;  119927 MW;  6A3FB3FC8C879A71 CRC64;

  Query Match             64.7%;  Score 4204.5;  DB 1;  Length 1119;
  Best Local Similarity   64.9%;  Pred. No. 1.2e-173;
  Matches  844;  Conservative  106;  Mismatches  128;  Indels  223;  Gaps   18;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              ||||:||| :|||:|:||||||||:|||||||||| | :|||| |||||||||||||||
Db          1 MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYNYKGPGTPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              |:|||||||||||||:|||||||||| |||||||| :|||| ||||||||:: |
Db         61 ASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGEEFMTFSTLTCTVSNTLTP 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              |||| ||||||:|||||||||| ||||||||||||||||||| | |||:|:||:| |||
Db        121 SIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTFNDGGKKISINVDFERSNVDP 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              ||  |||:||||||:|||||||| ||||||||||::: ||| ||||||||:|||||||
Db        181 KGYLTDSRVIPSLNKVSTLFVAPQCANGYTSGTMGFANTYGDVQIDCSNIHVGITKGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              |||||||||||||||:|||| | |:||||||||||:|||||||||| |||:| |:|||||
Db        241 WNYPVSSESFSYTKTCSSNGIFITYKNVPAGYRPFVDAYISATDVNSYTLSYANEYTCAG 300

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
                 |  |||||||||:|||||||||||||||||||||||||||:|: |||||||||:
Db        301 GYWQRAPFTLRWTGYRNSDAGSNGIVIVATTRTVTDSTTAVTTLPFDPNRDKTKTIEILK 360

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
              |||||||||||||||| ||||||||||||||:|||||||||:|||||:||| |||
Db        361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDIPYHTTTTVTSKWTGTITSTTTHTNP 420

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
              |||||||:|||| |||||:|||||||||||||||:| ||| ||||:|||||||||||||
Db        421 TDSIDTVIVQVPSPNPTVTTTEYWSQSFATTTTITGPPGNTDTVLIREPPNHTVTTTEYW 480

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
              |:|: ||:| |||||||||||:|||||||||||||:|: ||:| ||||||||||||:|
Db        481 SESYTTTSTFTAPPGGTDSVIIKEPPNPTVTTTEYWSESYTTTSTFTAPPGGTDSVIIKE 540

Qy        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
              ||| |||||||||||| ||||||||||||:|::|||||||||||||||||| ||||| |
Db        541 PPNHTVTTTEYWSQSYTTTTTVTAPPGGTDTVLVREPPNHTVTTTEYWSQSYTTTTTVIA 600

Qy        601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
              ||||||:|||||||| ||||||||||:|||||:| ||  ||||:|||||| ||||||||
Db        601 PPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYW 660

Qy        661 SQSYATTTTITAPPGETDTVL----------------------------------IRE 684
              |||||||||||||||||||||                                  |||
Db        661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDTVIIRE 720

Qy        685 PPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTA 744
              |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||: |
Db        721 PPNHTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTIIA 780

Qy        745 PPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSN 804
              ||| ||||:|||||||||||||||: | |||||||||||||||::||||:||
Db        781 PPGETDTVLIREPPNPTVTTTEYWSQSYTTATTVTAPPGGTDTVIIYDTMSSSEIS---- 836

Qy        805 DITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELIC 864
                    ||||||| | ||                          | ||| | |::| ||  |
Db        837 -------SFSRPHYTNHTT----------------------LWSTTWVIETKTITETSC 866

Qy        865 SDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGV 924
                | || :| |: ||| |:: |: :||:|| : :| | |  |  ||        |:|||
Db        867 EGDKGCSWVSVSTRIVTIPNNIETPMVTNTVDSTTTESTS-QSPSGI-------FSESGV 918

Qy        925 SVTTETS-VTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQST 983
              || ||:| ||| ||                  |||  :||:| ||: |:|::| ::| ||
Db        919 SVETESSTVTTAQTN----------------PSVPTTESEVVFTTKGNNENGPYESPST 961

Qy        984 STEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNG 1043
                    :                | |: |  || |
Db        962 N-----------------VKSSMDENSEFT---------------------------- 974

Qy       1044 DNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIE---ATT 1100
                        ||| ||                                  |||:||     |
Db        975 -------------TSTAAS----------------------------TSTDIENETIAT 992

Qy       1101 TSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSELVATTQATNEN-GGKSPSTD 1159
              |   || || :||  | |          ||::||::: | :    | || | |||:||
Db        993 TGSVEASSPIISSSADET----------TTVTTTAESTSVI---EQPTNNNGGGKAPSA- 1038

Qy       1160 LTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQ 1219
                    |:  | :|:||:: | :|        : | :|| | |  |     |:| ||||||
Db       1039 -----TSSPSTTTTANNDSVITG--------TTSTNQSQSQSQYN-----SDTQQTTLSQ 1080

Qy       1220 QVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
              |:|||  | :  : :|:|||||:||||||| ||||||||||
Db       1081 QMTSSLVSLH--MLTTFDGSGSVIQHSTWLCGLITLLSLFI 1119


RESULT 3
ALA1_CANAL
ID   ALA1_CANAL     STANDARD;      PRT;  1419 AA.
AC   O13368;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein ALA1 precursor (Agglutinin-like adhesin).
GN   ALA1 OR ALS5.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98053977; PubMed=9393828;
RA   Gaur N.K., Klotz S.A.;
RT   "Expression, cloning, and characterization of a Candida albicans
RT   gene, ALA1, that confers adherence properties upon Saccharomyces
```

```
RT   cerevisiae for extracellular matrix proteins.";
RL   Infect. Immun. 65:5289-5294(1997).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF025429; AAB88883.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18   1419       AGGLUTININ-LIKE PROTEIN ALA1.
FT   DOMAIN     399    404       POLY-THR.
FT   DOMAIN     408    418       POLY-THR.
FT   DOMAIN     437    441       POLY-THR.
FT   DOMAIN     673    676       POLY-SER.
FT   DOMAIN     687    690       POLY-SER.
FT   DOMAIN     700    703       POLY-SER.
FT   DOMAIN     719    724       POLY-SER.
FT   DOMAIN     749    752       POLY-SER.
FT   DOMAIN     787    791       POLY-SER.
FT   DOMAIN     869    872       POLY-SER.
FT   DOMAIN     875    883       POLY-SER.
FT   DOMAIN     901    911       POLY-SER.
FT   DOMAIN    1216   1221       POLY-SER.
FT   CARBOHYD   665    665       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   919    919       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1301   1301       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1326   1326       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE  1419 AA;  149635 MW;  249F33F688A9D5B6 CRC64;

  Query Match          49.2%;  Score 3194;  DB 1;  Length 1419;
  Best Local Similarity  49.5%;  Pred. No. 3.6e-130;
  Matches  708;  Conservative  182;  Mismatches  355;  Indels  186;  Gaps   29;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              |:|||||||||||| |:|| |||:|:| :|||||||| ||||||||||||||||||||||
Db          1 MIQQFTLLFLYLSFATAKAITGIFNSIDSLTWSNAGNYAFKGPGYPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              ||||||| |||||||:| || ||||||||||||||||||||||||:| |||| |:|
Db         61 ANPGDTFILNMPCVFKFTASQKSVDLTADGVKYATCQFYSGEEFTTFSSLKCTVNNNLRS 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              |||| |||||||||||||||| ||||||||||||||||||| | :|| | |||||||
Db        121 SIKALGTVTLPIAFNVGGTGSSVDLEDSKCFTAGTNTVTFNDGSKKLSIAVNFEKSTVDQ 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              | ||  || ||||||: ||:|||||||||||||||||:| |||||||||:||||:||:||
Db        181 SGYLTTSRFMPSLNKIATLYVAPQCENGYTSGTMGFSTSYGDVAIDCSNVHIGISKGVND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||:||:|||||||:|:| || | |||||||||||||||| :| ||| |:| ||||
Db        241 WNHPVTSESFSYTKSCSSFGISITYQNVPAGYRPFIDAYISPSDNNQYQLSYKNDYTCVD 300

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
              |  ||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DYWQHAPFTLKWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
              |||||||||||| |||| :||::||:|| :|||:|    |||:||:||| | ||||||:|
Db        421 TDSIDTVVVQVPSPNPTTTTTQFWSESFTSTTTITNSLKGTDSVIVREPHNPTVTTTEFW 480

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
              |:||||| |:|: | ||||||:||| |||||||:||:|:||| |:|  | ||||||:||
Db        481 SESFATTETITSKPEGTDSVIVREPHNPTVTTTEFWSESYATTETITNGPEGTDSVIVRE 540

Qy        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
              | ||||||:||:||||| |:|  | ||||||::|| | ||||||:||:||||| |:|
Db        541 PHNPTVTTTKFWSESYATTETITNKPEGTDSVIVKEPYNPTVTTTEFWSESYATTETITN 600

Qy        601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTT--- 657
              | |||:||:||| | ||||||:||:|:||| |:|  | |||:::| :|   : :||
Db        601 GPEGTDSVIVREPHNPTVTTTEFWSESYATTETITTGPLGTDSIVIHDPLEESSSTTAIE 660

Qy        658 ----------EYWSQSYATTTTITAPPGETDTVLIREPPNHTVT---------------- 691
                         :  | |   :::|      :|  |  : |: : |
Db        661 SSDSNISSSAQESSSSVEQSSSIVGLSSSSDIPLSSDMPSSSSTGLTSSESSTVSSYDSD 720

Qy        692 ----------------------TTEYWSQS----YATTTTVTAPPGETDTVLIREPPNHT 725
                                   || :|  |     :|: | ::      : |::   |
Db        721 SSSSIESSTLSSSDRCSSSISDTTSFWDSSSSDLESTSITWSSSIDAQSSHLVQSVSNSI 780

Qy        726 VTTTEYWSQSYATTTTVTAPPGGTDTVIIREP------------PNPTVTTTEYW----- 768
              |: |  | |   ::|       || ::  :             |: |:::::  :
Db        781 STSQELSSSSSEESSTF-----ATDALVSSDASSILSSDTSSYYPSSTISSSDDFPHTIA 835

Qy        769 ------SQSFATTT--------TVTAPPGGT-DTVIIYESMSSSKISTSSNDI------- 806
                   | || |:|        ::|: |  : |:     | |||  |:  :||
Db        836 GESDSLSISFITSTVEISSDSVSLTSDPASSFDSSSSLNSDSSSSPSSDQSDILTSSSFS 895

Qy        807 TSIIPSFS----------RPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTES 856
              | ::|||           ||||||||   |  |||| :|:  |        |  ::||
Db        896 TLVVPSFSLSSSSSLSLTYPHYVNSTTYHASESESSSVASPSMASESAN--DDTYTLSES 953

Qy        857 ETTTELICSDGKE---CSRLSSSSGIVTN-PDSNESSIVTSTVPTASTMSDSLSSTDG-I 911
                ||  | :|      | | :    |||  | |:  |  || | | |:   | : |
Db        954 TDTTSSIGTDSSTVTFCRRDNGDGCIVTGMPSSSIDSEQTSDVTTTSSFVASSTPTSAEQ 1013

Qy        912 SATSSDNV--SKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTS 969
              | | : |:  |::  | :|::||:   |  | :  | ||   |||  | | |:: :: |:
Db       1014 SITDNPNIDSSQTSASSSTKSSVSVSDTVVNSILLSETS--TLSSDDSTS-SDTSISSTT 1070

Qy        970 NGD----NQSGTHDSQSTSTEIEIVTTS---STKVLPPVVSSNTDLTSEPTNTR------ 1016
              | |    |   :| | ::  |  |  |    |:  |  :|| :|:| |   |
Db       1071 NSDTGNINAGSSHTSTASIKESSIQKTGVTLSSSYLSTKLSSTSDITIELITTELITTEL 1130

Qy       1017 -----EQPTTLSTTSNSITE------DITTSQPTGDNGDNTSSTNPVPTVAT----STLA 1061
                  :| | ::| :| :|       : : |   ::   || |  | ||::    || |
Db       1131 TTIEDNEPNTFTSTPSSHSEIFSSDNSVLSKQVDRESTIKTSPTTDVTTVSSLSVHSTEA 1190

Qy       1062 SA-----SEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTD 1116
              |      :   | : :  :|||: :   ::  | |:    :  |    |||  |  ||
Db       1191 STATLGENSFSNVASTPSNIATSLRSTSSSSNHATESSGTVKSEASAEAIPSPPTS--TD 1248

Qy       1117 VTTEPTDTREQPTTLSTTSKTN-----SELVATTQATN---ENGGKSPSTDLTSSLTTGT 1168
                :    :  | : :  ||      |:: | | :|:   ||   : : |  ||
Db       1249 NRLSYSTEEAKGITYANSGSTNNLITESQVAAPTDSTSVLIENPVVTSTFDDNSSAAVDQ 1308

Qy       1169 SASTSANSELV----TSGSVTGGAVASASNDQSHSTSV---------------TNSNSI 1208
                : | :  | :    ::      | :|: |  |  |:|:                | :|
Db       1309 PSKTKSIEESIMNPDSTNETNNGFIATLSQAQVPSSSIHSELISTTTAKTTDASMNGDSA 1368

Qy       1209 VSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLF 1259
                ||:  ||| ||| :|  | |  : :|| || |  :| :||   |::   |
Db       1369 ASNSQPTTLIQQVATS--SYNQPLITTYAGSSSATKHPSWLLKFISVALFF 1417


RESULT 4
ALS2_CANAL
ID   ALS2_CANAL     STANDARD;      PRT;   468 AA.
AC   O74657;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein 2 precursor (Fragment).
GN   ALS2.
```

```
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98440424; PubMed=9765564;
RA   Hoyer L.L., Payne T.L., Hecht J.E.;
RT   "Identification of Candida albicans ALS2 and ALS4 and localization of
RT   als proteins to the fungal cell surface.";
RL   J. Bacteriol. 180:5334-5343(1998).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF024580; AAC64235.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   >468       AGGLUTININ-LIKE PROTEIN 2.
FT   CARBOHYD    253    253       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    315    315       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     468    468
SQ   SEQUENCE   468 AA;  50203 MW;  BFE773E169ED0FAF CRC64;

  Query Match             30.7%;  Score 1992.5;  DB 1;  Length 468;
  Best Local Similarity   80.2%;  Pred. No. 3.5e-79;
  Matches  376;  Conservative   33;  Mismatches   59;  Indels    1;  Gaps    1;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              || || || | :|:|:|| |||||:||:|||| | |||:||| ||||||||||||||||
Db          1 MLLQFLLLSLCVSVATAKVITGVFNSFDSLTWTRAGNYAYKGPNRPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              |||||||||||||||: | |||||||:|||||||||||||||||:| |||:: | |
Db         61 ANPGDTFTLNMPCVFKFITDQTSVDLTAEGVKYATCQFYSGEEFTTFSSLKCTVSNTLTS 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              |||| |||||||:|||||||| ||| |:|| |||||||||||| |||||:|||:  |
Db        121 SIKALGTVTLPISFNVGGTGSSVDLESSQCFKAGTNTVTFNDGDKKISIDVDFEKTNEDA 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              | |  |||::||:|||:  :||||| |||||| |||    ||  |||||:|:||||||||
Db        181 SGYFIASRLIPSINKVSITYVAPQCANGYTSGAMGFIVLTGDTTIDCSNVHVGITKGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||:||||:| || |||:| || | |:||||||||| | | | :  |:  | ||||| | |
Db        241 WNFPVSSDSLSYNKTCSSTGISITYENVPAGYRPFFDVYTSVSGQNR-QLRYTNDYACVG 299

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
              | |||||| ||  || ||:| ||| ||||||||||||||||||||||||||||||||||||
Db        300 SSLQSKPFNLRLRGYNNSEANSNGFVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 359

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        360 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 419

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREP 469
              |||||||||||| |||||:||||||||:|||||||||||||||:||||||
Db        420 TDSIDTVVVQVPSPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREP 468


RESULT 5
ALS4_CANAL
ID   ALS4_CANAL     STANDARD;      PRT;   469 AA.
AC   O74660;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Agglutinin-like protein 4 precursor (Fragment).
GN   ALS4.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98440424; PubMed=9765564;
RA   Hoyer L.L., Payne T.L., Hecht J.E.;
RT   "Identification of Candida albicans ALS2 and ALS4 and localization of
RT   als proteins to the fungal cell surface.";
RL   J. Bacteriol. 180:5334-5343(1998).
CC   -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
CC   -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF024584; AAC64239.1; -.
KW   Cell adhesion; Glycoprotein; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   >469       AGGLUTININ-LIKE PROTEIN 4.
FT   NON_TER     469    469
SQ   SEQUENCE   469 AA;  49604 MW;  0EDCAB19B89EFCB1 CRC64;

  Query Match             28.0%;  Score 1819;  DB 1;  Length 469;
  Best Local Similarity   72.1%;  Pred. No. 9.8e-72;
  Matches  338;  Conservative   48;  Mismatches   83;  Indels    0;  Gaps    0;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              || || || | :|:|:|| |||||:|||||||:|||:| ::||  ||| ||:|||||| :
Db          1 MLLQFLLLSLCVSVATAKVITGVFNSFNSLTWANAASYPYRGPATPTWTAVIGWSLDGAT 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              |: ||||||:||||||: | |||:|| |||  ||||   | |||||||:::|||   : :
Db         61 ASAGDTFTLDMPCVFKFITDQTSIDLVADGRTYATCNLNSAEEFTTFSSVSCTVTTTMTA 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
                || |||||| :|:|||:|| || :|:||||| |||||||||  ||  |:||||||
Db        121 DTKAIGTVTLPFSFSVGGSGSDVDLANSQCFTAGINTVTFNDGDTSISTTVDFEKSTVAS 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              |  :  ||::|||::   ||:  :| ||||||||||::     |||| :|:||: ||||
Db        181 SDRILLSRILPSLSQAVNLFLPQECANGYTSGTMGFSTAGTGATIDCSTVHVGISNGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||||:|||||||||||||  : : :||||||||||:||||||| |: ||: ||| | | |
Db        241 WNYPISSESFSYTKTCTSTSVLVTFQNVPAGYRPFVDAYISATRVSSYTMQYTNIYACVG 300

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
              :      ||  | || || |||||| || ||||||||||||||||   ||||||||||
Db        301 AASVDDSFTHTWRGYSNSQAGSNGITIVVTTRTVTDSTTAVTTLPFNSDTDKTKTIEILQ 360

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREP 469
              |||||||||||| |||||:||||||||:|||||||||||||||:||||||
Db        421 TDSIDTVVVQVPSPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREP 469


RESULT 6
```

```
FLO1_YEAST
ID   FLO1_YEAST     STANDARD;      PRT;  1537 AA.
AC   P32768;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Flocculation protein FLO1 precursor (Flocculin 1).
GN   FLO1 OR YAR050W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94262325; PubMed=8203162;
RA   Watari J., Takata Y., Ogawa M., Sahara H., Koshino S., Onnela M.-L.,
RA   Airaksinen U., Jaatinen R., Penttilae M., Keraenen S.;
RT   "Molecular cloning and analysis of the yeast flocculation gene FLO1.";
RL   Yeast 10:211-225(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=95249563; PubMed=7731988;
RA   Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA   Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA   Storms R.K.;
RT   "The nucleotide sequence of chromosome I from Saccharomyces
RT   cerevisiae.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
RN   [3]
RP   PRELIMINARY SEQUENCE FROM N.A.
RX   MEDLINE=93289821; PubMed=8511970;
RA   Teunissen A.W.R.H., Holub E., van der Hucht J., van den Berg J.A.,
RA   Steensma H.Y.;
RT   "Sequence of the open reading frame of the FLO1 gene from
RT   Saccharomyces cerevisiae.";
RL   Yeast 9:423-427(1993).
RN   [4]
RP   REVIEW.
RX   MEDLINE=96076625; PubMed=7502576;
RA   Teunissen A.W., Steensma H.Y.;
RT   "Review: the dominant flocculation genes of Saccharomyces cerevisiae
RT   constitute a new subtelomeric gene family.";
RL   Yeast 11:1001-1013(1995).
CC   -!- FUNCTION: MAY BE DIRECTLY INVOLVED IN THE FLOCCULATION PROCESS.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC       (Potential).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE FLOCCULIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X78160; CAA55024.1; -.
DR   EMBL; L28920; AAC09499.1; ALT_SEQ.
DR   PIR; S31230; S31230.
DR   SGD; S0000084; FLO1.
DR   InterPro; IPR001389; Flocculin.
DR   Pfam; PF00624; Flocculin; 18.
KW   Glycoprotein; Membrane; Repeat; Cell wall; Signal; GPI-anchor.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25      ?       FLOCCULATION PROTEIN FLO1.
FT   PROPEP        ?   1537       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN      274   1240       SER/THR-RICH.
FT   DOMAIN      278   1087       18 X 45 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT      278    322       1-1.
FT   REPEAT      323    367       1-2.
FT   REPEAT      368    412       1-3.
FT   REPEAT      413    457       1-4.
FT   REPEAT      458    502       1-5.
FT   REPEAT      503    547       1-6.
FT   REPEAT      548    592       1-7.
FT   REPEAT      593    637       1-8.
FT   REPEAT      638    682       1-9.
FT   REPEAT      683    727       1-10.
FT   REPEAT      728    772       1-11.
FT   REPEAT      773    817       1-12.
FT   REPEAT      818    862       1-13.
FT   REPEAT      863    907       1-14.
FT   REPEAT      908    952       1-15.
FT   REPEAT      953    997       1-16.
FT   REPEAT      998   1042       1-17.
FT   REPEAT     1043   1087       1-18.
FT   DOMAIN     1118   1157       2 X 20 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT     1118   1137       2-1.
FT   REPEAT     1138   1157       2-2.
FT   DOMAIN     1226   1392       3 X 51 AA APPROXIMATE REPEATS.
FT   REPEAT     1226   1276       3-1.
FT   REPEAT     1291   1341       3-2.
FT   REPEAT     1342   1392       3-3.
FT   DOMAIN     1408   1434       3 X 9 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT     1408   1416       4-1.
FT   REPEAT     1417   1425       4-2.
FT   REPEAT     1426   1434       4-3.
FT   CARBOHYD    135    135       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    262    262       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    329    329       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    374    374       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    419    419       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    509    509       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    554    554       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    599    599       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    644    644       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    689    689       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    734    734       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1114   1114       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1537 AA;  160722 MW;  992650C6BE9A8CEA CRC64;

  Query Match             14.5%;  Score 940;  DB 1;  Length 1537;
  Best Local Similarity   26.0%;  Pred. No. 1.9e-33;
  Matches  417;  Conservative  244;  Mismatches  523;  Indels  418;  Gaps   85;

Qy          2 LQQFTLLFLYLSIASAKT---------ITGVFDSF------NSLTWSNAANYAF------ 40
              |  |||| |  |:|| |          :|:  :|      :| |:|||| |:
Db         10 LAVFTLLAL-TSVASGATEACLPAGQRKSGMNINFYQYSLKDSSTYSNAAYMAYGYASKT 68

Qy         41 ----------------------------KGPGYPTW----------NAVLGWSLD--G 58
                                          :  | |            : || | |
Db         69 KLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGACSNSQGIAYWSTDLFG 128

Qy         59 TSANPGDTFTLNM----------PCVFKYTTSQ----------------------TSVD 85
                | :  || |            ||: |                          || :
Db        129 FYTTPTNV-TLEMTGYFLPPQTGSYTFKFATVDDSAILSVGGATAFNCCAQQQPPITSTN 187

Qy         86 LTADGVK----------YATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGT----VTLP 131
                | ||:|           |   |:|  :           : |:  ::||    ||||
Db        188 FTIDGIKPWGGSLPPNIEGTVYMYAGYYYPM---------KVVYSNAVSWGTLPISVTLP 238

Qy        132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV-DPSAYLYASRVM 190
                 ||  | | |           :|:|       |:     || ||| |   |
Db        239 -----DGTTVSDDFEG--------YVYSFDD-------DLSQSNCTVPDPSNYA-VSTTT 277

Qy        191 PSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESF 250
               :    | | :   |    :|| | :    : |          |      ||   :| |
Db        278 TTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTTASTIITTTEPWNSTFTSTST 337

Qy        251 SYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTL 310
               |    :||::     :       |    :||     |  : : :|   : |
Db        338 ELTTVTGTNGVRTDETII------VIRTPTTATTAITTTEPWNSTFTSTSTEL------- 384
```

```
Qy       311 RWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVTTL-PFNPSVDKTKT---------- 355
Db       385 ------TTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGL 438

Qy       356 ------IEILQPIPTTTITT-------SYVGVTTSYLTKTA----PIGETATVIVDVPYH 398
Db       439 PTDETIIVIRTPTTATTAMTTTQPWDDTFTSTSTEMTTVTGTNGLPTDET-IIVIRTPTT 497

Qy       399 TTT--TVTSEWTGTIT------TTTTRTN--PTDSIDTVVVQVPLPNPT-VSTTEYWSQS 447
Db       498 ATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDE-TIIVIRTPTTATTAMTTTQPWNDT 556

Qy       448 FATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTVTAPP 494
Db       557 FTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTEPWNSTFTSTSTELTTVTGTN 616

Qy       495 G-GTDS--VIIREPPNPT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPP 542
Db       617 GLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPT 676

Qy       543 NPT--VTTTEYWSQSYATT----TTVTAPPG-GTDS--VIIREPPNHT--VTTTEYWSQS 591
Db       677 TATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDT 736

Qy       592 YATT----TTVTAPPG-GTD--TVIIREPPNHTV--TTTEYWSQSFATT----TTVT--- 635
Db       737 FTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTVTGTN 796

Qy       636 GPPSGTDTVIIREPPNP--TVTTTEYWSQSYATT----TTITAP---PGETDTVLIREPP 686
Db       797 GQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMTTITGTNGVPTDETVIVIRTPT 856

Qy       687 NHTV--TTTEYWSQSYATT----TTVTAPPGE-TD--TVLIREPPNHTV--TTTEYWSQS 735
Db       857 SEGLISTTTEPWTGTFTSTSTEMTTITGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGT 916

Qy       736 YATT----TTVTAPPG-GTD--TVIIREPPNPTV--TTTEYWSQSFATTTTVTAPPGGT- 785
Db       917 FTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEVTTITGTN 976

Qy       786 -----DTVIIYESMSSSKI------------STSSNDITSIIPSFSRP---------- 816
Db       977 GQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPT 1036

Qy       817 --HYVNSTTSD-LSTFESSSMNTPTSISSDGMLLSSTTLVTESETT----TELICSDGKE 869
Db      1037 SEGLVTTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSSSGQI 1096

Qy       870 CSRLSSSSGIVTNP--DSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVT 927
Db      1097 TSSITSSRPIIT-PFYPSNGTSVISSSVISSSVTSSLFTSSPVISSSVISSSTTTSTSIF 1155

Qy       928 TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEI 987
Db      1156 SESSKSSV--IPTSSSTSGSSESETSSAGSVSSS----SFISSESSKSPTYSSSS----L 1205

Qy       988 EIVTTSST-----KVLPPVVSSNTDLTSEPTNTREQPTTLSTTS-------NSITEDITT 1035
Db      1206 PLVTSATTSQETASSLPPATT---------TKTSEQTTLVTVTSCESHVCTESISPAIVS 1256

Qy      1036 SQPTGDNGDNTSSTN--PVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTS 1093
Db      1257 TATVTVSGVTTEYTTWCPISTTETTKQTKGTTEQTTETTKQTTVVTISSCESDVCSKTAS 1316

Qy      1094 TEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTT-------SKTNSELVATT- 1145
Db      1317 PAIVSTSTATINGVTTEYTTWCPIST--TESRQQTTLVTVTSCESGVCSETASPAIVSTA 1374

Qy      1146 ---------------QATNENGGKSPSTDLTSSLTTGT-SASTSANSELVTSGSVTGGA 1188
Db      1375 TATVNDVVTVYPTWRPQTANEESVSSKMNSATGETTTNTLAAETTTNTVAAETITNTGAA 1434

Qy      1189 ------------------VASASNDQSHSTSV-----TNSNSIVSNTPQTTLSQQVTSS 1224
```

```
Db      1435 ETKTVVTSSLSRSNHAETQTASATDVIGHSSSVVSVSETGNTKSLTSSGLSTMSQQPRST 1494

Qy      1225 SP-----STNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db      1495 PASSMVGYSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1536
```

RESULT 7
YAG3_YEAST
ID YAG3_YEAST STANDARD; PRT; 1322 AA.
AC P39712;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Hypothetical 138.1 kDa protein in FLO9-GDH3 intergenic precursor.
GN YAL063C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC -!- SIMILARITY: STRONG, TO YEAST PROTEIN FLO1.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U12980; AAC04971.1; -.
DR SGD; S0000059; YAL063C.
DR InterPro; IPR001389; Flocculin.
DR Pfam; PF00624; Flocculin; 13.
KW Hypothetical protein; Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1322 HYPOTHETICAL PROTEIN YAL063C.
FT TRANSMEM 366 388 POTENTIAL.
FT TRANSMEM 754 775 POTENTIAL.
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 257 257 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 270 270 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 464 464 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 554 554 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 599 599 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 644 644 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 689 689 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 734 734 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 888 888 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1322 AA; 138072 MW; AADFD1FF13267CEA CRC64;

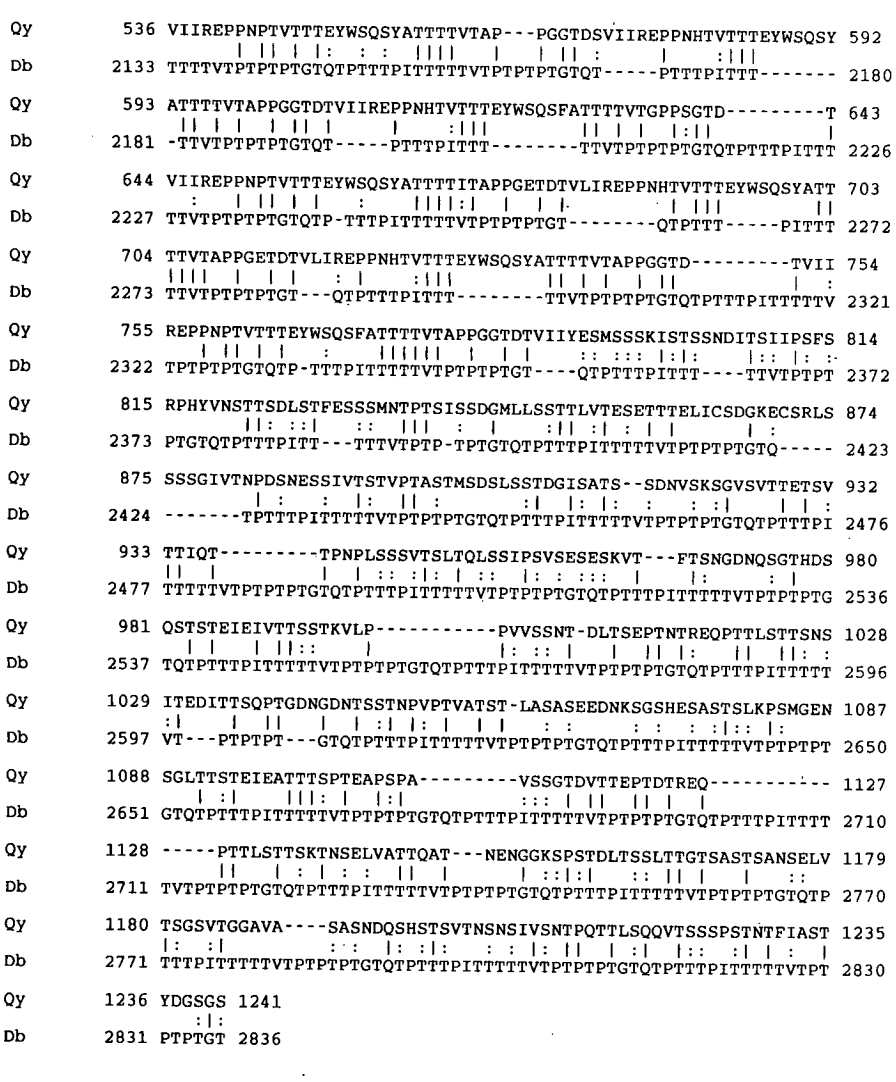
Query Match 14.4%; Score 932.5; DB 1; Length 1322;
Best Local Similarity 28.1%; Pred. No. 3.3e-33;
Matches 400; Conservative 215; Mismatches 468; Indels 339; Gaps 79;

```
Qy        28 NSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKYTTSQT----S 83
```

```
Db        50 DSSTYSNAAYMAY---GYASKT-----KLGSVGGQTDISIDYNIPCV---SSSGTFPCPQ 98

Qy        84 VDLTAD-GVK-YATCQ------FYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIA-- 133

Db        99 EDLYGNWGCKGIGACSNNPIIAYWSTDLFGFYTTPT--------------NVTLEMTGY 143

Qy       134 FNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSL 193

Db       144 FLPPQTGSYT-----------------------------FKFATVDDSAIL--------- 165

Qy       194 NKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDC----------SNIHIGITKGLNDWN- 242

Db       166 ------------------------SVGGSIAFECCAQEQPPITSTNFTI---NGIKPWNG 198

Qy       243 -YPVSSESFSYTKTCTSNGIQIKYQNVPA-GYRPF-------------IDAYISATDVN- 286

Db       199 SPPDNITGTVYMYAGFYYPMKIVYSNAVAWGTLPISVTLPDGTTVSDDFEGYVYTFDNNL 258

Qy       287 ---QYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDA-------GSNGI-----VIVATT 331

Db       259 SQPNCTIPDPSNYTVSTTITTTEP----WTGTFTSTSTEMTTVTGTNGVPTDETVIVIRT 314

Qy       332 RTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTIT-TSYVGVTTSYLTKTAPIGETAT 390

Db       315 PT-TASTIITTTEPWNSTFTSTST-------ELTTVTGTNGVRTDETIIVIRTPTTATTA 366

Qy       391 VIVDVPYHTTTTVTSEWTGTITTTTTRTN--PTDSIDTVVVQVPLPNPT-VSTTEYWSQS 447

Db       367 ITTTEPWNSTFTSTS----TELTTVTGTNGLPTDE-TIIVIRTPTTATTAMTTTQPWNDT 421

Qy       448 FATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTVTAPP 494

Db       422 FTSTSTELTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTVTGTN 481

Qy       495 G-GTDS--VIIREPPNPT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPP 542

Db       482 GLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTDETIIVIRTPT 541

Qy       543 NPT--VTTTEYWSQSYATT----TTVTAPPG-GTDS--VIIREPPNHT--VTTTEYWSQS 591

Db       542 TATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTEPWNST 601

Qy       592 YATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTVTGP- 637

Db       602 FTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTQPWNDTFTSTSTEMTTVTGTN 661

Qy       638 --PSGTDTVIIREPPNPT--VTTTEYWSQSYATT----TTITAP---PGETDTVLIREPP 686

Db       662 GLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTDETIIVIRTPT 721

Qy       687 NHT--VTTTEYWSQSYATT----TTVTAP---PGETDTVLIREPPNHTV--TTTEYWSQS 735

Db       722 TATTAMTTTQPWNDTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGT 781

Qy       736 YATT----TTVTAPPG-GTD--TVIIREPPNP--TVTTTEYWSQSFATT----TTVTAPP 782

Db       782 FTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMTTITGTN 841

Qy       783 G--GTDTVIIYESMS---SSKISTSSNDITSIIPS---FSRPHYVNSTTSDLSTFESSSM 834

Db       842 GQPTDETVIIVKTPTTAISSSLSSSSGQITSFITSARPIITPFYPSNGTSVISSSVISSS 901

Qy       835 NTPTSISSDGM---LLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIV 891

Db       902 DTSSLVISSSVTSSLVTSSPVISSSFISSPVISSTTTSASILSESS------------K 948

Qy       892 TSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQ 951

Db       949 SSVIPTSSSTSGSSESETGSASSAS---SSSSISSESPKSTYSSSSLP-PVTSATTSQEI 1004

Qy       952 LSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSE 1011

Db      1005 TSSLPPVTTTKT-----------------SEQTTLVTVTSCESHVCTESISSAIVSTAT 1046

Qy      1012 PT---NTREQPTTLSTTSNSITEDIT-TSQPTGDNGDNTSSTNPVPTVATSTLASASEED 1067

Db      1047 VTVSGATTEYTTWCPISTTEITKQTTETTKQTKGTTEQTTETTKQTTVVT---ISSCESD 1103

Qy      1068 --NKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTR 1125

Db      1104 VCSKTASPAIVSTSTATINGVTTEYTTWCPISTTESKQQTTLVTVTSCGSGVCSETT--- 1160

Qy      1126 EQPTTLSTTSKTNSELVAT-----TQATNENGGKSPSTDLTSSLTTGTSAST------SA 1174

Db      1161 -SPAIVSTATATVNDVVTVYSTWRPQTTNEQSVSSKMNSATSETTTNTGAAETTTSTGAA 1219

Qy      1175 NSELVTSGSVT-----GGAVASASNDQSHSTSV-----TNSNSIVSNTPQTTLSQQVTSS 1224

Db      1220 ETKTVVTSSISRFNHAETQTASATDVIGHSSSVVSVSETGNTKSLTSSGLSTMSQQPRST 1279

Qy      1225 ------SPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260

Db      1280 PASSMVGSSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1321
```

RESULT 8
MUC2_HUMAN
ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN MUC2 OR SMUC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor.";
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace.R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=1885763;
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -!- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A
CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS
CC AGENTS AT MUCOSAL SURFACES.
CC -!- SUBUNIT: MULTIMERIC.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
CC BRONCHUS, CERVIX AND GALL BLADDER.
CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND

```
CC       VARIES AMONG DIFFERENT ALLELES.
CC   -!- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
CC       OF SILKWORM HEMOCYTIN.
CC   -!- SIMILARITY: CONTAINS 2 VWFC DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE (CTCK) DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L21998; AAB95295.1; -.
DR   EMBL; M74027; AAA59875.1; -.
DR   EMBL; M94131; AAA59163.1; -.
DR   EMBL; M94132; AAA59164.1; -.
DR   Genew; HGNC:7512; MUC2.
DR   MIM; 158370; -.
DR   InterPro; IPR000359; Cys_knot.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR002400; GF_cysknot.
DR   InterPro; IPR002919; TIL_Cysrich.
DR   InterPro; IPR001007; VWF_C.
DR   InterPro; IPR001846; VWF_D.
DR   Pfam; PF00007; Cys_knot; 1.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF00094; vwd; 4.
DR   Pfam; PF01826; TIL; 1.
DR   PRINTS; PR00438; GFCYSKNOT.
DR   SMART; SM00214; VWC; 2.
DR   SMART; SM00216; VWD; 4.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 2.
KW   Glycoprotein; Repeat; Signal.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21   5179       MUCIN 2.
FT   DOMAIN    1401   1747       APPROXIMATE REPEATS.
FT   REPEAT    1401   1416       1.
FT   REPEAT    1417   1432       2.
FT   REPEAT    1433   1448       3.
FT   REPEAT    1449   1464       4.
FT   REPEAT    1465   1471       5.
FT   REPEAT    1472   1478       6.
FT   REPEAT    1479   1494       7A.
FT   REPEAT    1495   1517       7B.
FT   REPEAT    1518   1533       8A.
FT   REPEAT    1534   1556       8B.
FT   REPEAT    1557   1572       9A.
FT   REPEAT    1573   1596       9B.
FT   REPEAT    1597   1612       10A.
FT   REPEAT    1613   1635       10B.
FT   REPEAT    1636   1651       11A.
FT   REPEAT    1652   1675       11B.
FT   REPEAT    1676   1683       12.
FT   REPEAT    1684   1699       13.
FT   REPEAT    1700   1715       14.
FT   REPEAT    1716   1731       15.
FT   REPEAT    1732   1747       16..
FT   DOMAIN    4815   4886       VWFC 1.
FT   DOMAIN    4924   4991       VWFC 2.
FT   DOMAIN    5075   5160       CTCK.
FT   DISULFID  5075   5122       BY SIMILARITY.
FT   DISULFID  5089   5136       BY SIMILARITY.
FT   DISULFID  5098   5152       BY SIMILARITY.
FT   DISULFID  5102   5154       BY SIMILARITY.
FT   DISULFID     ?   5159       BY SIMILARITY.
FT   CARBOHYD   163    163       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   423    423       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   670    670       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   770    770       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   894    894       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1139   1139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1154   1154       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1215   1215       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1230   1230       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1246   1246       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1787   1787       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1820   1820       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4339   4339       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4351   4351       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4362   4362       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4373   4373       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4422   4422       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4438   4438       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4502   4502       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4616   4616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4627   4627       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4752   4752       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4787   4787       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4881   4881       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4888   4888       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4955   4955       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  4970   4970       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  5019   5019       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  5038   5038       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  5069   5069       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT  1351   1351       H -> L (IN REF. 3).
FT   CONFLICT  1412   1412       T -> S (IN REF. 3).
FT   CONFLICT  1449   1449       L -> P (IN REF. 3).
FT   CONFLICT  1504   1504       M -> T (IN REF. 3).
FT   CONFLICT  4192   4192       G -> S (IN REF. 2).
SQ   SEQUENCE   5179 AA;  540295 MW;  85CD7571FB9A5663 CRC64;

  Query Match           11.1%;  Score 721.5;  DB 1;  Length 5179;
  Best Local Similarity  25.3%;  Pred. No. 1.7e-23;
  Matches  335;  Conservative  158;  Mismatches  554;  Indels  279;  Gaps   54;

Qy        59 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 118
             |:  |  | | : |     ||:  |  :|       |         |  | :| | :
Db      1647 TTTTPPPTTTPSPP-----TTTTPSPPITTTTTPPPTT--------TPSSPITTTPSPPT 1693

Qy       119 KSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV 178
             :     | |         | |||       | | : |        ::      :|:
Db      1694 TTMTTPSPTTTPSSPITTTTTPSSTTTPSPPPTTMTTPSPTTTPSPPTTTM------TTL 1747

Qy       179 DP---SAYLYASRVMPSLNKVT---------TLFVAPQCE-NGY-TSGTMGFSSSNGDVA 224
              |   |: |  : : ||:   |         |    | |   |:  ||   |    ||
Db      1748 PPTTTSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVPLCNWTGWLDSGKPNFHKPGGDTE 1807

Qy       225 IDCSNIHIGITKGLNDWNYPVSSESFSY----------TKTC-TSNGIQIKYQN-VPAGY 272
             :      ||   |   |   :|  :  |          |  |  | |:  | ::  | |
Db      1808 L------IGDVCG-PGWAANISCRATMYPDVPIGQLGQTVVCDVSVGLICKNEDQKPGGV 1860

Qy       273 RPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTR 332
              |             : | |  :  |      ::| |:  |  :|    :   :   ||
Db      1861 IPM----------AFCLNYEINVQCC--ECVTQPTTMTTTTTENPTPPTTTPITTTTTV 1907

Qy       333 TVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIG-ETATV 391
             | | : |   |    |    |       |  | | ||: :  ||:    | | :| |
Db      1908 TPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTT 1967

Qy       392 IVDVPYHTTTTVTSEWTGT---------ITTTT----------TRTNPTDSIDTVVVQVP 432
                 | ||||||   | |         |||||          |:|  |  | |     |
Db      1968 ---TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTP 2024

Qy       433 LPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTA 492
             | || : |   :    ||||||  | | |   : |    :|||        || | |
Db      2025 TPTPTGTQTP-TTTPITTTTTVTPTPTPTGT---QTPTTTPITTT--------TTVTPTP 2072

Qy       493 PPGGTD-------SVIIREPPNPTVTTTEYWSQSFATTTTVTAP---PGGTD-------S 535
             | ||          :      | || | |:  : :  ||||   |   | ||         :
Db      2073 TPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPIT 2132
```



```
Qy        536 VIIREPPNPTVTTTEYWSQSYATTTTVTAP---PGGTDSVIIREPPNHTVTTTEYWSQSY 592
Db       2133 TTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQT-----PTTTPITTT------- 2180

Qy        593 ATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTD---------T 643
Db       2181 -TTVTPTPTPTGTQT-----PTTTPITTT--------TTVTPTPTPTGTQTPTTTPITTT 2226

Qy        644 VIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATT 703
Db       2227 TTVTPTPTPTGTQTP-TTTPITTTTTTVTPTPTPTGT--------QTPTTT-----PITTT 2272

Qy        704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTD---------TVII 754
Db       2273 TTVTPTPTPTGT---QTPTTTPITTT--------TTVTPTPTPTGTQTPTTTPITTTTTV 2321

Qy        755 REPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFS 814
Db       2322 TPTPTPTGTQTP-TTTPITTTTTTVTPTPTPTGT----QTPTTTPITTT----TTVTPTPT 2372

Qy        815 RPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLS 874
Db       2373 PTGTQTPTTTPITT---TTTVTPTP-TPTGTQTPTTTPITTTTTTVTPTPTPTGTQ----- 2423

Qy        875 SSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATS--SDNVSKSGVSVTTETSV 932
Db       2424 -------TPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPI 2476

Qy        933 TTIQT---------TPNPLSSSVTSLTQLSSIPSVSESESKVT---FTSNGDNQSGTHDS 980
Db       2477 TTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTG 2536

Qy        981 QSTSTEIEIVTTSSTKVLP----------PVVSSNT-DLTSEPTNTREQPTTLSTTSNS 1028
Db       2537 TQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTT 2596

Qy       1029 ITEDITTSQPTGDNGDNTSSTNPVPTVATST-LASASEEDNKSGSHESASTSLKPSMGEN 1087
Db       2597 VT---PTPTPT---GTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPT 2650

Qy       1088 SGLTTSTEIEATTTSPTEAPSPA---------VSSGTDVTTEPTDTREQ---------- 1127
Db       2651 GTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTT 2710

Qy       1128 -----PTTLSTTSKTNSELVATTQAT---NENGGKSPSTDLTSSLTTGTSASTSANSELV 1179
Db       2711 TVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTP 2770

Qy       1180 TSGSVTGGAVA----SASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIAST 1235
Db       2771 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPT 2830

Qy       1236 YDGSGS 1241
Db       2831 PTPTGT 2836
```

```
RESULT 9
FLO5_YEAST
ID   FLO5_YEAST     STANDARD;      PRT;  1075 AA.
AC   P38894;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Flocculation protein FLO5 precursor (Flocculin 5).
GN   FLO5 OR YHR211W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
RN   [2]
RP   REVIEW.
RX   MEDLINE=96076625; PubMed=7502576;
RA   Teunissen A.W., Steensma H.Y.;
RT   "Review: the dominant flocculation genes of Saccharomyces cerevisiae
RT   constitute a new subtelomeric gene family.";
RL   Yeast 11:1001-1013(1995).
CC   -!- FUNCTION: MAY BE DIRECTLY INVOLVED IN THE FLOCCULATION PROCESS.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC       (Potential).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED (PROBABLE).
CC   -!- SIMILARITY: BELONGS TO THE FLOCCULIN FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U00029; AAB69731.1; -.
DR   PIR; S48992; S48992.
DR   SGD; S0001254; FLO5.
DR   InterPro; IPR001389; Flocculin.
DR   Pfam; PF00624; Flocculin; 8.
KW   Glycoprotein; Membrane; Repeat; Cell wall; Signal; GPI-anchor.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25      ?       FLOCCULATION PROTEIN FLO5.
FT   PROPEP        ?   1075       REMOVED IN MATURE FORM (POTENTIAL).
FT   CARBOHYD    135    135       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    187    187       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    203    203       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    262    262       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    663    663       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    749    749       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1075 AA;  111981 MW;  D151B370B60C8D9F CRC64;

  Query Match           10.9%;  Score 711;  DB 1;  Length 1075;
  Best Local Similarity   25.9%;  Pred. No. 8.3e-24;
  Matches  324;  Conservative  192;  Mismatches  393;  Indels  342;  Gaps   66;

Qy        153 AGTNTVTFNDGDKDISIDVEF------EKSTVDPSAYL---YASRV---------- 189
Db         23 SGATEACLPAGQRKSGMNINFYQYSLKDSSTYSNAAYMAYGYASKTKLGSVGGQTDISID 82

Qy        190 --MPSLNKVTTLFVAPQCENGYTS-GTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVS 246
Db         83 YNIPCVSSSGT-FPCPQ-EDSYGNWGCKGMGA--------CSN-----SQGIAYW----S 123

Qy        247 SESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSR---L 303
Db        124 TDLFGFYTTPTNVTLEMTGYFLPPQTGSYTFSFATVDDSAILSVGGSIAFECCAQEQPPI 183

Qy        304 QSKPFTLRWTGYKNSDAGSNGIVIVATT----------RTVTDSTTAVTTLPFNPSVDKT 353
Db        184 TSTNFTI--NGIKPWD-GSLPDNITGTVYMYAGYYYPLKVVYSNAVSWGTLPI--SVE-- 236

Qy        354 KTIEILQPIPT-TTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTT---TTVTSEWTG 409
Db        237 --------LPDGTTVSDNFEGYVYSF---DDDLSQSNCTIPDPSIHTTSTITTTTEPWTG 285
```

```
Qy        410 TIT------TTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDT 463
              | |      || | ||   : :||:|   :  ||            | :|:|
Db        286 TFTSTSTEMTTITDTNGQLTDETVIV---IRTPT-----------TASTIT-------- 322

Qy        464 VIIREPPNHTVTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPPNP--TVTTTE 514
                  |||| |: :| :|    ||||   |  ||   ::|| | :     ||||
Db        323 ----------TTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLITTTTE 371

Qy        515 YWSQSFATT----TTVTAPPG-GTDS--VIIREPPNP--TVTTTEYWSQSYATTTTVTAP 565
              |: :| :|    ||||   |  ||  ::|| | :     |||| |: :: :|:|
Db        372 PWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEVTT 431

Qy        566 PGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNH--TVTTTEY 623
                ||:     :| : ||                          ::|| | :     ||||
Db        432 ITGTNG----QPTDETV------------------------IVIRTPTSEGLITTTTEP 462

Qy        624 WSQSFATT----TTVT---GPPSGTDTVIIREPPNPTV--TTTEYWSQSYATT----TTI 670
              |: :| :|    ||||   | |:    ::|| | :   :  |||| |: :: :|    |||
Db        463 WTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEVTTI 522

Qy        671 TAPPGE-TD--TVLIREPPNH--TVTTTEYWSQSYATT----TTVTAPPGE-TD--TVLI 718
              |   |: ||   ::|| | :     |||| |: :: :|    ||||   |: ||   ::|
Db        523 TGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVI 582

Qy        719 REPPNH--TVTTTEYWSQSYATTTTVTAPPGGT------DTVIIREPPNPTVTTTEYWSQ 770
              | | :      ||| |: :: :|:|      ||      :|||:   |  |  ::   |
Db        583 RTPTSEGLITRTTEPWTGTFTSTSTEVTTITGTNGQPTDETVIVIRTPT-TAISSSLSSS 641

Qy        771 SFATTTTVTA---------PPGGTDTV---IIYESMSSSKISTSSNDITSIIPSFSRPHY 818
              |   |:::|:         |  ||  :   :|  |::|| :::||   :|:| |
Db        642 SGQITSSITSSRPIITPFYPSNGTSVISSSVISSSVTSSLVTSSSFISSSVISS------ 695

Qy        819 VNSTTSDLSTFESSSMNT--PTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSS 876
                |||:  | |  || ::  ||| |: |    | |:  : | ::: :     |  :  |||
Db        696 --STTTSTSIFSESSTSSVIPTSSSTSGSSESKTSSASSSSSSSSISSESPKSPTNSSSS 753

Qy        877 SGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATS------SDNVSKSGVSVTTET 930
                ||:  : : :   |::| |:|   |   |  :: ||      ::::| : ||  | |
Db        754 LPPVTSATTGQET--ASSLPPATTTKTS-EQTTLVTVTSCESHVCTESISSAIVSTATVT 810

Qy        931 --SVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIE 988
                 |||  ||  |:|:: |:                        || :    :||
Db        811 VSGVTTEYTTWCPISTTETT----------------------KQTKGTTEQTKGTTEQT 847

Qy        989 IVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSS 1048
                ||  | |:  : |  :|: |:  :    |  :||:: :|            ||  |
Db        848 TETTKQTTVV-TISSCESDICSKTAS----PAIVSTSTATI-----------NGVTTEY 890

Qy       1049 TNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAP- 1107
              |   |   | :                          : : | | |  |:   | | :|
Db        891 TTWCPISTTES----------------------KQQTTLVTVTSCESGVCSETTSPA 925

Qy       1108 --SPAVSSGTDVTT-----EPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDL 1160
                 | | ::  || |      |  | ||    | :|| ||   ||:: |   |     | :
Db        926 IVSTATATVNDVVTVYPTWRPQTTNEQ----SVSSKMNS---ATSETTTNTGAAETKTAV 978

Qy       1161 TSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSV-----TNSNSIVSNTPQT 1215
              ||||:    | |                 |||::   ||:||     | :   ::::  :
Db        979 TSSLSRFNHAETQ--------------TASATDVIGHSSSVVSVSETGNTMSLTSSGLS 1023

Qy       1216 TLSQQVTSS------SPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
              |:|||  |:        || :   ||| || : :   : |   |  | | |
Db       1024 TMSQQPRSTPASSMVGSSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1074
```

```
RESULT 10
AMYH_YEAST
ID   AMYH_YEAST     STANDARD;      PRT;  1367 AA.
AC   P08640; P08068;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
DE   glucosidase) (1,4-alpha-D-glucan glucohydrolase).
GN   STA1 OR STA2 OR MAL5 OR YIR019C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RA   Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA   Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA   Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA   Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA   Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA   Walsh S.V., Whitehead S.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX   MEDLINE=87194600; PubMed=3106330;
RA   Yamashita I., Nakamura M., Fukui S.;
RT   "Gene fusion is a possible mechanism underlying the evolution of
RT   STA1.";
RL   J. Bacteriol. 169:2142-2149(1987).
RN   [3]
RP   SEQUENCE OF 1-31 FROM N.A.
RC   STRAIN=SPX101-1C;
RX   MEDLINE=89031230; PubMed=3141213;
RA   Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT   "Similar short elements in the 5' regions of the STA2 and SGA genes
RT   from Saccharomyces cerevisiae.";
RL   FEBS Lett. 239:179-184(1988).
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC       glucose residues successively from non-reducing ends of the chains
CC       with release of beta-D-glucose.
CC   -!- SIMILARITY: TO S.POMBE SPBC215.13.
CC   -!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z38061; CAA86176.1; -.
DR   EMBL; M16164; AAA35014.1; -.
DR   EMBL; M16165; AAA35015.1; -.
DR   EMBL; X13857; CAA32069.1; -.
DR   PIR; B26877; B26877.
DR   PIR; A26877; A26877.
DR   PIR; S48478; S48478.
DR   SGD; S0001458; MUC1.
KW   Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW   Signal; Multigene family.
FT   SIGNAL       1     21       POTENTIAL.
FT   CHAIN       22   1367       GLUCOAMYLASE S1/S2.
FT   DOMAIN     210   1367       SER/THR-RICH.
FT   CARBOHYD   817    817       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   874    874       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1367 AA;  136110 MW;  91C00E2DBD61AA9D CRC64;

  Query Match            10.4%;  Score 675.5;  DB 1;  Length 1367;
  Best Local Similarity   26.2%;  Pred. No. 3.6e-22;
  Matches  378;  Conservative  177;  Mismatches  535;  Indels  353;  Gaps   65;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              | : | | :| ||:           |||                |:||     | | :|||
Db          1 MQRPFLLAYLVLSLL----------FNSAL-----------GFPTALVPRG-SSEGTS 36

Qy         61 AN------PGDTFTLNM--PCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTC 112
                 |      |   |  :|    : :||   |||    |   |       |:|
```

```
Db        37 CNSIVNGCPNLDFNWHMDQQNIMQYTLDVTSVSWVQDNT-YQITIHVKGKE--------- 86

Qy       113 TVNDALKS--SIKAFGTVTLPIAFNVGGTGSSTDLEDSKC-FTAGTNTVTFN-DGDKDI- 167

Db        87 --NIDLKYLWSLKIIGVTGPKGTVQLYGYNENTYLIDNPTDFTA-----TFEVYATQDVN 139

Qy       168 -------SIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSN 220

Db       140 SCQVWMPNFQIQFE--------YLQGS--------------AAQYASSWQWGTTSFDLST 177

Qy       221 GDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYI 280

Db       178 G-----CNNYD---NQGHSQTDFP----GFYWNIDCDNN--------------------C 205

Qy       281 SATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA 340

Db       206 GGTKSSTTTSSTSESSTTTSSTSESSTTT-------SSTSESS----TTTSSTSESSTSS 254

Qy       341 VTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTT 400

Db       255 STTAPATPT-----TTSCTKEKPTPPTTTS--------CTKEKP---------TPPHHDT 292

Qy       401 TTVTSEWTGTITTTTTRTN---PTDSIDTV-VVQVPLPNPTVSTTEYWSQ--SFATTTTV 454

Db       293 TPCTKKKTTTSKTCTKKTTTPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESS 352

Qy       455 TAP---PGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVT 511

Db       353 SAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAP-----------VPTPSSS 401

Qy       512 TTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTT-VTAPPGGTD 570

Db       402 TTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVPTP 461

Qy       571 SVIIREPPNHTVT--TTEYWS-----QSYATTTTVTAPPGGTDTVIIREP-PNHTVTTTE 622

Db       462 SSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTE 521

Qy       623 YWS-----QSFATTTTVTGPPSGTDTVIIREP-PNPTVTTTEYWSQSYATTTTITAPPGE 676

Db       522 SSSAPAPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTE------SSSTPVTSSTTE 575

Qy       677 TDTVLIREPPNHTVTTTEYWSQSYAT----TTTVTAPPGETDTVLIREPPNHTVT--TTE 730

Db       576 SSSAPV---PTPSSSTTESSSAPVPTPSSSTTESSSAPAPTPSSSTTESSSAPVTSSTTE 632

Qy       731 YWS-----QSYATTTTVTAP---PGGTDTVIIREP-PNPTVTTTEYWSQSFATTTTVTAP 781

Db       633 SSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTE--SSSAPVTSSTTES 690

Qy       782 PGGTDTVIIYESMSS-----SKISTSSNDITSIIPSFSR--------PHYVNSTTSD--- 825

Db       691 SSAPVTSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSA 750

Qy       826 ---LSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTN 882

Db       751 PVTSSTTESSSAPVPTP-SSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSVAPV 809

Qy       883 PDSNESSIVTSTVPTASTMSDSL-SSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNP 941

Db       810 PTPSSSSNITSSAPSSTPFSSSTESSSVPVPTPSSSTTESSSAPVSSSTTESSVAPVPTP 869

Qy       942 LSSSVTSLTQLSSIPSVSESES-----KVTFTSNGDNQSGTHDSQSTSTEIEIV------ 990

Db       870 SSSSNITSSAPSSIPFSSTTESFSTGTTVTPSSSKYPGSQTETSVSSTTETTIVPTKTTT 929

Qy       991 ---TTSSTKVLPPVVSSNTDLTSEPTN-------TREQPTTLST--TSNSITEDITTSQP 1038

Db       930 SVTTPSTTTITTTVCSTGTNSAGETTSGCSPKTVTTTVPTTTTTSVTTSSTTTITTTVCS 989

Qy      1039 TGDN--GDNTSSTNPVPTVATSTLASASEEDNKSGSHESA-----------------ST 1078

Db       990 TGTNSAGETTSGCSP-KTITTTVPCSTSPSETASESTTTSPTTPVTTVVSTTVVTTEYST 1048

Qy      1079 SLKPSMGENSGLTTS--TEIEATTTSPTEAPSPAV----------------SSGTD---- 1116

Db      1049 STKPG-GE---ITTTFVTKNIPTTYLTTIAPTPSVTTVTNFTPTTITTTVCSTGTNSAGE 1104

Qy      1117 ----------VTTEPTDT-----REQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLT 1161

Db      1105 TTSGCSPKTVTTTVPCSTGTGEYTTEATTLVTTAVTTT-VVTTESSTGTNSAGKTTTGYT 1163

Qy      1162 SSLTTGTSASTSANSELVTSG------SVTGGAVASASNDQSHSTSVTNSNSIVSN---- 1211

Db      1164 TKSVPTTYVTTLAPSAPVTPATNAVPTTITTTECSAATNAAGETTSVCSAKTIVSSASAG 1223

Qy      1212 --------TPQTT---LSQQVTSSSPSTN----------------TFIASTYDGSGSIIQ 1244

Db      1224 ENTAPSATTPVTTAIPTTVITTESSVGTNSAGETTTGYTTKSIPTTYITTLIPGSNGAKN 1283

Qy      1245 HST 1247

Db      1284 YET 1286
```

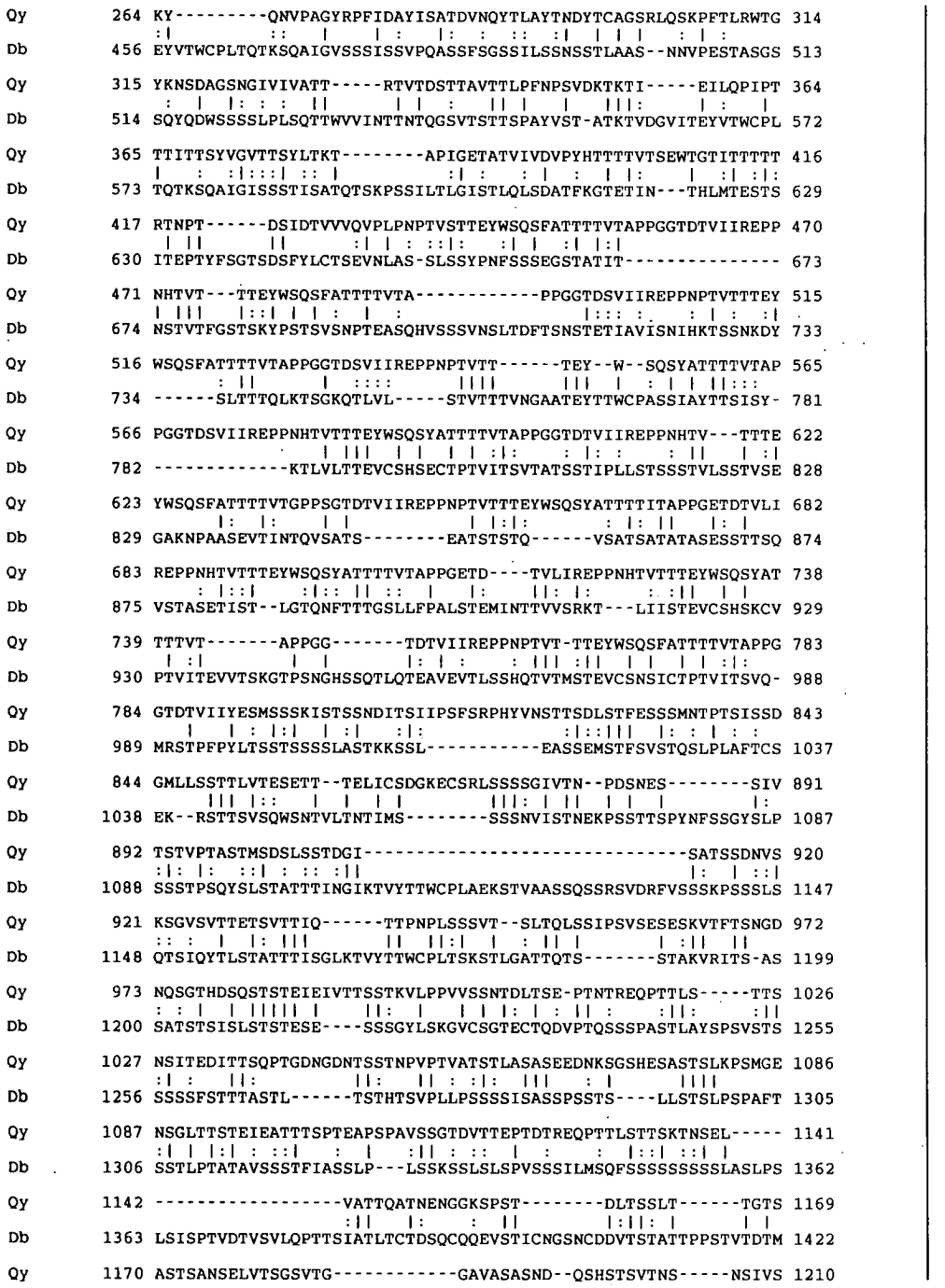
RESULT 11
DAN4_YEAST
ID   DAN4_YEAST   STANDARD;   PRT; 1161 AA.
AC   P47179;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cell wall protein DAN4 precursor.
GN   DAN4 OR YJR151C OR J2223.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Scarcez T.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REGULATION.
RX   MEDLINE=21113168; PubMed=11160904;
RA   Cohen B.D., Sertil O., Abramova N.E., Davies K.J., Lowry C.V.;
RT   "Induction and repression of DAN1 and the family of anaerobic
RT   mannoprotein genes in Saccharomyces cerevisiae occurs through a
RT   complex array of regulatory sites.";
RL   Nucleic Acids Res. 29:799-808(2001).
CC   -!- FUNCTION: COMPONENT OF THE CELL WALL (By similarity).
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC   (Potential).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE SRP1 / TIP1 FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; Z49651; CAA89684.1; -.
DR   SGD; S0003912; DAN4.
DR   InterPro; IPR000992; SRP1_TIP1.
DR   Pfam; PF00660; SRP1_TIP1; 1.
DR   PROSITE; PS00724; SRP1_TIP1; 1.
KW   Cell wall; Glycoprotein; Membrane; GPI-anchor; Signal.
FT   SIGNAL   1   24   POTENTIAL.
FT   CHAIN   25   1146   CELL WALL PROTEIN DAN4.
FT   PROPEP   1147   1161   REMOVED IN MATURE FORM (POTENTIAL).
FT   LIPID   1146   1146   GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   1161 AA; 118358 MW; 7954C15D69F0CA58 CRC64;

```
 Query Match             8.9%;  Score 577;  DB 1;  Length 1161;
 Best Local Similarity   26.3%;  Pred. No. 5e-18;
 Matches  293;  Conservative  131;  Mismatches  427;  Indels  264;  Gaps   46;

Qy        276 IDAYISATDVNQYTLAY--------TNDY--TCAGSRLQSKPFTLRWTGYKNSDAGSNGI 325
              |:  :  :|:  :   |        |  |    | :      || | ||    :
Db         36 IELAVYVSDIRAHIFQYYSFRNHHKTETYPSEIAAAVFDYGDFTTRLTGISGDE------ 89

Qy        326 VIVATTRTVTDSTTAVTTLPFNPSVDKTKTIE-ILQPIPTTTITTSYVGVTTSYLTKTAP 384
                || :|      | |   |::    : : |   |||:| ||     ||   | | |
Db         90 ----VTRMITGVPWYSTRL--KPAISSALSKDGIYTAIPTSTSTT-----TTKSSTSTTP 138

Qy        385 IGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYW 444
                        |||:||      ||:|| | || |  :          || |||
Db        139 --------------TTTITS------TTSTTSTTPTTSTTST-------TPTTSTT--- 167

Qy        445 SQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 504
              | :  |:|| | |   | :          |  ||   | :  |:|| | |   | |
Db        168 STTPTTSTTSTTPTTSTTSTTPTTSTTSTTPTTSTTSTTPTTSTTSTTPTT 227

Qy        505 PPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTA 564
                || :||   ||:   :|| |      |           || :||          |:|| ||
Db        228 STTPTTSTTSTTSQTSTKSTTPTTSSTST---------TPTTSTTP------TTSTTSTA 272

Qy        565 PPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 624
              |   | |          | :||   : | | ||: |:    | :       :  ::||
Db        273 PTTSTTS---------TTSTTS--TISTAPTTSTTSSTFSTSSA----SASSVISTTATT 317

Qy        625 SQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIRE 684
              | :||: ||      : ||          :|:||  :: |  ||||       :|| :
Db        318 STTFASLTTPATSTASTDHT------TSSVSTTNAFTTSATTTTT-------SDTYISSS 364

Qy        685 PPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQ---------- 734
              |:   :: |  : |  |::    |   :      ||     | :|: |
Db        365 SPSQVTSSAEPTTVSEVTSS--VEPTRSSQVTSSAEP----TTVSEFTSSVEPTRSSQVT 418

Qy        735 SYATTTTV---TAPPGGTDTVIIREPPNPTVTTTEYWSQ----------SFATTTTVTAP 781
              | |  |||   |:    | :  :     || | :|: |           | |  |||:
Db        419 SSAEPTTVSEFTSSVEPTRSSQVTSSAEPT-TVSEFTSSVEPTRSSQVTSSAEPTTVSEF 477

Qy        782 PGGTDTVIIYESMSSSKISTSSNDITSIIP-------SFSRPHYVNSTTSDLSTFESSSM 834
                  :     : ||:: :| |   :|: |       | : |  |:  || :    || :
Db        478 TSSVEPTRSSQVTSSAEPTTVSEFTSSVEPIRSSQVTSSAEPTTVSEVTSSVEPIRSSQV 537

Qy        835 NTPTSISSDGMLLSSTTLVTE----SETTT-------------ELICSDGKECSRLSSS 876
              |   :|| |   |  |   |    |: ||             ||| |     | ::||
Db        538 TTTEPVSSFGSTFSEITSSAEPLSFSKATTSAESISSNQITISSELIVS-----SVITSS 592

Qy        877 SGIVTNPDSNESSIVTSTV-PTA---STMSDSLSSTDGISATS----------------- 915
              | | :: :   || ::|:| ||:    :  :|:|||: :||||
Db        593 SEIPSSIEVLTSSGISSSVEPTSLVGPSSDESISSTESLSATSTFTSAVVSSSKAADFFT 652

Qy        916 -SDNVSKSGVSVTTETSVTTIQTTPN-PLSSSVTSLTQ----------LSSIPSVSESES 963
               |   :|| ||  : |  ||   ||: ||: | | :|           | | |  || :
Db        653 RSTVSAKSDVSGNSSTQSTTFFATPSTPLAVSSTVVTSSTDSVSPNIPFSEISSSPESST 712

Qy        964 KVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNT-----REQ 1018
               :| ||       |     :|: :   | |:  |   :|||    : |||::
Db        713 AITSTSTSFIAERTSSLYLSSSNMSSFTLSTFTVSQSIVSS---FSMEPTSSVASFASSS 769

Qy       1019 PTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTV-----ATSTLASASEEDNKSGSH 1073
              |  :|: ||    |  :|        || |:    |:     ||||  : | :
Db        770 PLLVSSRSN--CSDARSS--------NTISSGLFSTIENVRNATSTFTNLSTD------- 812

Qy       1074 ESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTT--EPTDTREQPTTL 1131
              |   || | |       : | |::    |: |         ||   ||   |  |
Db        813 EIVITSCKSSCTNEDSVLTKTQVSTVETTITSC------SGGICTTLMSPVTTINAKANT 866

Qy       1132 STTSKTNSELVATTQATNENGGKS----PSTDLTSSLTTGTSASTSANSELVTS------ 1181
               ||::|::  | ||  | . |  |     | | :||  ||  : |   | | :||
Db        867 LTTTETST--VETTITTCPGGVCSTLTVPVTTITSEATTTATISCEDNEEDITSTETELL 924

Qy       1182 ------GSVTGGAVASASN-----DQSHSTSVTNSNSIVSNTPQT-------TLSQQVTS 1223

                    | :||   :  :     :   :|  |   | |  |  |       ||:  ||:
Db        925 TLETTITSCSGGICTTLMSPVTTINAKANTLTTTETSTVETTIITTCSGGVCSTLTVPVTT 984

Qy       1224 -SSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLS 1257
               :| :|  |   |  |     :    | |  : | ::
Db        985 ITSEATTTATISCEDNEEDVASTKTELLTMETTIT 1019
```

```
RESULT 12
YS89_CAEEL
ID   YS89_CAEEL     STANDARD;      PRT;  3178 AA.
AC   Q09624; Q09625; Q969D4;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein ZK945.9 in chromosome II.
GN   ZK945.9/ZK945.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Wilkinson-Sproat J.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REVISIONS.
RA   Durbin R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z48544; CAB70192.1; -.
DR   EMBL; Z48582; CAB70192.1; JOINED.
DR   EMBL; Z48582; CAB70201.1; -.
DR   EMBL; Z48544; CAB70201.1; JOINED.
DR   WormPep; ZK945.9; CE25697.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR001024; Lipoxygenase_LH2.
DR   InterPro; IPR000636; M+channel_nlg.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00520; ion_trans; 1.
DR   Pfam; PF01477; PLAT; 1.
DR   Pfam; PF01825; GPS; 1.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00308; LH2; 1.
KW   Hypothetical protein; Transmembrane.
FT   DOMAIN      266   1196       SER/THR-RICH.
FT   DOMAIN     1105   1241       GLY/SER-RICH.
FT   DOMAIN     2071   2120       GPS.
FT   TRANSMEM     13     30       POTENTIAL.
FT   TRANSMEM     51     73       POTENTIAL.
FT   TRANSMEM   2139   2161       POTENTIAL.
FT   TRANSMEM   2348   2367       POTENTIAL.
FT   TRANSMEM   2390   2412       POTENTIAL.
FT   TRANSMEM   2451   2468       POTENTIAL.
FT   TRANSMEM   2483   2505       POTENTIAL.
FT   TRANSMEM   2567   2589       POTENTIAL.
FT   TRANSMEM   2836   2858       POTENTIAL.
FT   TRANSMEM   2939   2961       POTENTIAL.
FT   TRANSMEM   2976   2998       POTENTIAL.
FT   TRANSMEM   3038   3060       POTENTIAL.
SQ   SEQUENCE   3178 AA;  344726 MW;  F8239436D03666CD CRC64;
```

```
  Query Match             8.5%;  Score 553.5;  DB 1;  Length 3178;
  Best Local Similarity   24.3%;  Pred. No. 1.6e-16;
  Matches  290;  Conservative  154;  Mismatches  401;  Indels  349;  Gaps   51;

Qy        252 YTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLR 311
              | : | |  |  :|       ||:      :| ::     | :     | |  :| |::|
Db        194 YREKCESGEINEEYARRMC-KRPYRSE--KSTAISDSQGVYYDGQVLKGVR--AKQFSMR 248

Qy        312 WTG-----YKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTT 366
              :|       ||| |      :||:  || |   :|  | |              |||
Db        249 TSGSPTLRRMKRDAGDN------TCDYTIESTSTSTTTPTTTTVTSTVT-------STTT 295

Qy        367 ITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 426
              : ||   |||:  | |:    : |:       |:|| ||    | :|:|: |: |    :
Db        296 VPTSTSTVTTAMSTSTSTPSTSTTI-----ESTSTTFTS----TASTSTSSTSTTQQSSS 346

Qy        427 VVVQVPLPNPTVSTTEYWSQSFATTTT--VTAPPGGTDTVIIREPPNHTVTTTEYWSQSF 484
              :  |  : |:||      |  ||||  :|:         |    : | |:| : :
Db        347 TITSSP-SSTTLST------SIPTTTTPEITSTLSSLPDNAICSYLDETTTSTTFTTTML 399

Qy        485 ATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNP 544
              :|||              | |: : ||||  | |   |||                ||
Db        400 TSTTT-------------EEPSTSTTTTEVTSTSSTVTTT-------------EPTTT 431

Qy        545 TVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGG 604
              |:|   | :  :|:|||  |          |   |||:      | :::|||| |
Db        432 LTTSTASTSTTEPSTSTVTTSPS-------TSPVTSTVTS------SSSSSTTVTTP--- 475

Qy        605 TDTVIIREPPNHTVTTTEYWSQSFATTTTVTGP------PSGTDTVIIREPPNPTVTTTE 658
              | |      |: ||||:     : | :|: |||      || | :  :    : | ::|
Db        476 TSTESTSTSPSSTVTTS-----TTAPSTSTTGPSSSSSTPSSTASSSVSSTASSTQSSTS 530

Qy        659 YWSQSYATTT---TITAPPGETDTVLIREPPNHTVTTTEYWSQS-------------YAT 702
              :|  :|||   | |:  |          |      ||| | | |              | |
Db        531 --TQQSSTTTKSETTTSSDGTNPDFYFVE----KATTTFYDSTSVNLTLNSGLGIIGYQT 584

Qy        703 TTTVTAPPGETDTVLIRE-------------------------PPNHTVTTT--------- 729
              :  |:|          ::                       | |:|   |
Db        585 SIECTSPTSSNYVSTTKDGACFTKSVSMPRLGGTYPASTFVGPGNYTFRATMTTDDKKVY 644

Qy        730 ----EYWSQSYATTT-----TVTAPPGGTDTVIIREPPNP-----TVTTTEYWSQSFATT 775
                  : | |::||      : :|    | :     |  |      | | ||  |   : :
Db        645 YTYANVYIQEYSSTTIESESSTSAVASSTSST----PSTPSSTLSTSTVTEPSSTRSSDS 700

Qy        776 TTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMN 835
              || :|  | | |  : || ::|: ||: :  |:|  : :      :|||:| ::  : |::
Db        701 TTTSA--GSTTT--LQESTTTSEESTTDSSTTTISDTSTTSSSPSSTTADSTS--TLSVD 754

Qy        836 TPTSISSDGMLLSST----------TLVTESETTTELICSDGKECSRLSS---------- 875
                |   |:  : |           | | : | ||   |   ||  :|:
Db        755 QFDFILDSGLSWNETRHNEDSINIVPLPTNAITPTER--SQTFECRNVSTEPFLIIKEST 812

Qy        876 ------------SSGIVTNP-----------------------------------DSN 886
                         || |   |                                    ||
Db        813 CLNYSNTVLNATYSSNIPIQPIETFLVGIGTYEFRINMTDLTTMQVVSHIFTLNVVADST 872

Qy        887 ESSIVTSTVPT-ASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSS 945
              :| ||||  | :|: | ::|:| || :||:   | :  |  : || :   ||     |
Db        873 STSEVTSTTSTGSSSESSAISTTSGIESTSTLEASTTDASQDSSTSTSDSGTT-----SD 927

Qy        946 VTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTST------------EIEIVTT 992
              |::   :| || |:|      :    : |   ||  |:|             :| :|
Db        928 STTIDSSNSTPSTSDSSG---LSQTPSDSSSASDSMRTTTVDPDASTETPYDFVLENLTW 984

Qy        993 SSTK-------VLPPVVSS-----NTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTG 1040
              : |         : |: :       | :| :  |   ||  |   || :||       |
Db        985 NETVYYSENPFYITPIPNKEPGALTTAMTCQCRNDSSQPFVLLKESNCLTE-------FG 1037

Qy       1041 DNGDNTSST--NPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGEN-SGLTTSTEIE 1097
              ||  ::|   ||: :   :|          |::|       : ||: | : |   :
Db       1038 KNGAYSASVSFNPMTSFVPAT----------GTYEFLINVTNRASGESASHIFTMNVVL 1086

Qy       1098 ATTTSPTEAPSPAVSS---------GTDVT--TEPTDTREQPTTLSTTSKTNSELVATTQ 1146
```

```
                |||  || |   |||          ||  |  |  | :    |||||      : | :|
Db       1087 PTTT--TETPPTTVSSSDDAGGKTGGTGATGGTGGTGSGGSATTLST-----GDAVRSTT 1139

Qy       1147 ATNENGGKSPSTDLTSSLTT----GTSASTSANSELVTSGSVT---------------- 1185
              : : :|  |  :    | ||    |:  |:   |: | ||  |
Db       1140 SGSGSGQSSTGSGAGGSGTTASGSGSGGSSGTGSDGVNSGKTTALNGDGTGSGTATTPGS 1199

Qy       1186 ----GGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQV-----TSSSPSTNT 1230
                  ||: : :| : |: |:  :|  |:|  :: |         :: ||| |
Db       1200 HLGDGGSTSGSGSDSNGSSGVSTKSSSGSDTSGSSDSSGANGAFSATAQPSTRT 1253


RESULT 13
FIG2_YEAST
ID   FIG2_YEAST     STANDARD;      PRT;  1609 AA.
AC   P25653;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Factor induced gene 2.
GN   FIG2 OR YCR089W OR YCR89W OR YCR1102.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92397594; PubMed=1523889;
RA   Wilson C., Grisanti P., Frontali L.;
RT   "The complete sequence of a 6146 bp fragment of Saccharomyces
RT   cerevisiae chromosome III contains two new open reading frames.";
RL   Yeast 8:569-575(1992).
CC   -!- FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC   -!- INDUCTION: BY MATING PHEROMONES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59720; CAA42254.1; -.
DR   PIR; S19504; S19504.
DR   PIR; S25345; S25345.
DR   SGD; S0000685; FIG2.
SQ   SEQUENCE   1609 AA;  166049 MW;  7D66AD7F85A7B852 CRC64;

  Query Match             8.4%;  Score 547.5;  DB 1;  Length 1609;
  Best Local Similarity   22.5%;  Pred. No. 1.3e-16;
  Matches  332;  Conservative  232;  Mismatches  503;  Indels  411;  Gaps   66;

Qy         55 SLDGTSANPGDTFT-LNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCT 113
              ||  || ||  ::|  |     | :   :::| |:  :  :|    | | | |  |:| :
Db        165 SLTFTSVNPSQSWTSFNSE---KSSALSSTIDFTSSEISGSTSP-KSLESFDTTGTITSS 220

Qy        114 VNDA-------------------------LKSSIKAF----GTVTLPIAFNVGGTGSST 143
                : :                       | |:|:|      : |:|   :   :   |
Db        221 YSPSPSSKNSNQTSLLSPLEPLSSSSGDLILSSTIQATTNDQTSKTIPTLVDATSSLPPT 280

Qy        144 DLEDSKCFTAGTNTVTFN--------DGDKDISIDVEFEKSTVDPSAYL----YASRVMP 191
                |   |:|::::: |         |: |:        :::|||  :     |:|  :
Db        281 LRSSSMAPTSGSDSISHNFTSPPSKTSGNYDV-----LTSNSIDPSLFTTTSEYSSTQLS 335

Qy        192 SLNKVT--------------------TLFVAPQCENGYT-SGTMGFSSSNGDVAIDCSN 229
              |||: :                    |  : |    | | |  :| |::|      : :
Db        336 SLNRASKSETVNFTASIASTPFGTDSATSLIDPISSVGSTASSFVGISTANFSTQGNSNY 395

Qy        230 IHIGITKG---LNDW---NYPVS-------------------SESFSYTKTCTSNGIQI 263
              :       |     ||   : |:|                 | ::  | | | :|:
Db        396 VPESTASGSSQYQDWSSSSLPLSQTTWVVINTTNTQGSVTSTTSPAYVSTATKTVDGVIT 455
```



RESULT 14
YM96_YEAST
ID YM96_YEAST STANDARD; PRT; 1140 AA.
AC Q04893;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 113.1 kDa protein in PRE5-FET4 intergenic region.
GN YMR317W OR YM9924.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC -!- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; Z54141; CAA90835.1; -.
DR SGD; S0004936; YMR317W.
KW Hypothetical protein; Repeat.
SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

Query Match 8.0%; Score 521.5; DB 1; Length 1140;
Best Local Similarity 24.4%; Pred. No. 1.2e-15;
Matches 245; Conservative 194; Mismatches 381; Indels 186; Gaps 40;

```
Qy       720 EPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVT 779
Db       347 EAPSSTSSSVSSEISSTTSSSVSSEAPLATSSVVSSEAPSSTSSSVSSEAPSSTSSSVSS 406

Qy       780 APPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTS 839
Db       407 EAPSSTSSSVSSE-ISSTKSSVMSSE------------VSSATSSLVSSEAPS--AISS 450

Qy       840 ISSDGMLLSSTTLVTESETTTEL--ICSDGKECSRLSSSSGIVTNPDSN--ESSIVTSTV 895
Db       451 LASSRLFSSKNTSVTSTLVATEASSVTSSLRPSSETLASNSIIESSLSTGYNSTVSTTTS 510

Qy       896 PTASTMSDSLSSTDGISATS-----SDNVSKSGVSVTTETSVTT-----IQTTPNPLSSS 945
Db       511 AASSTLGSKVSSSNSRMATSKTSSTSSDLSKSSVIFGNSSTVTTSPSASISLTASPLPSV 570

Qy       946 VTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIE-IVTTSSTKVLPPVVSS 1004
Db       571 WSDITS-------SEASS---ISSNLASSSAPSDNNSTIASASLIVTKTKNSVVSSIVSS 620

Qy      1005 NTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASAS 1064
Db       621 IT--SSETTN----ESNLATSSTSLLSNKATARSLSTS--NATSASNVPTGTFSSMSSHT 672

Qy      1065 EEDNKSGSHESASTSLKPSMGENS--GLTTSTEIEATTTS---PTEAPSPAVSSGTDVTT 1119
Db       673 SVITPGFSTSSASLAINSTVVSSSLAGYSFSTPESSPTTSTLVTSEAPS-TVSSMT--TS 729

Qy      1120 EP-----TDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTTGTSA---- 1170
Db       730 APFINNSTSARPSPSTASFITESTSSISSVPLASGDVTSSLAAHNLTTFSAPSTSSAQLV 789

Qy      1171 --STSANSELVT-----SGSVTGGA------------VASASNDQSHSTSVTNSNSIVSN 1211
Db       790 SKSTTSSSILVTPRIDRSGNSSTASRIATSLPNKTTFVSSLSSTSAHARNIFNSTVLATA 849

Qy      1212 TPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLS 1257
Db       850 KQIETLTSTVNCSNPTPNYNITKTV-----IVSRETTAIGTVTSCS 890


RESULT 15
YK82_YEAST
ID   YK82_YEAST     STANDARD;      PRT;  1169 AA.
AC   P36170;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   Hypothetical 122.2 kDa protein in SIR1 3'region precursor.
GN   YKR102W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Gaillon L., Dujon B.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: CELL WALL (BY SIMILARITY).
CC   -!- PTM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY).
CC   -!- SIMILARITY: STRONG, TO YEAST PROTEIN FLO1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z28327; CAA82182.1; -.
DR   PIR; S38181; S38181.
DR   SGD; S0001810; YKR102W.
KW   Hypothetical protein; Glycoprotein; Repeat; Cell wall; Signal.
FT   SIGNAL       1     26       POTENTIAL.
FT   CHAIN       27   1169       HYPOTHETICAL PROTEIN YKR102W.
FT   DOMAIN     274    609       SER/THR-RICH.
FT   CARBOHYD   122    122       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   157    157       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   279    279       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   389    389       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   452    452       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   656    656       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   686    686       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   879    879       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1092   1092       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1099   1099       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1169 AA;  122164 MW;  940C6DFB3569C669 CRC64;

  Query Match             7.7%;  Score 501;  DB 1;  Length 1169;
  Best Local Similarity   23.1%;  Pred. No. 9.2e-15;
  Matches  313;  Conservative  208;  Mismatches  528;  Indels  308;  Gaps   53;

Qy         6 TLLFLYLSIASAKTIT-----------GVFDSFNSLTWSNAANYAFKGPGYPTWNAVLG 53
Db        11 TGLFL-LSVANVALGTTEACLPAGEKKNGMTINFYQYSLKDSSTYS--NPSYMAYGYADA 67

Qy        54 WSLDGTSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGE-------EFTT 106
Db        68 EKLGSVSGQTKLSIDYSIPCNGASDTCACSDD---DATEYSASQVVPVKRGVKLCSDNTT 124

Qy       107 FSTLT-------CTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVT 159
Db       125 LSSKTEKRENDDCDQGAAYWSS-DLFGFYTTPTNVTVEMTG--------YFLPPKTGTYT 175

Qy       160 FNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSS 219
Db       176 FG------------FATVDDSAIL--------------------------------SV 189

Qy       220 NGDVAIDCSNIH---IGIT----KGLNDWN--YPVSSESFSYTKTCTSNGIQIKYQN-VP 269
Db       190 GGNVAFECCKQEQPPITSTDFTINGIKPWNADAPTDIKGSTYMYAGYYYPIKIVYSNAVS 249

Qy       270 AGYRPFIDAYISATDVNQ----YTLAYTND--------------YTCAGSRLQSKPFTL 310
Db       250 WGTLPVSVVLPDGTEVNDDFEGYVFSFDDNATQAHCSVPNPAEHARTCVSSATSS----- 304

Qy       311 RWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTS 370
Db       305 -WSSSE-----------VCTECTETESTSYVTPYVTSSSWSSSEVCTECTETESTSTSTP 352

Qy       371 YVGVTTSYLTKT-APIGETATVIVDVPYHTTTTVTSEWTGTITTTT---TRTNPTDSIDT 426
Db       353 YVTSSSSSSSEVCTECTETESTSYVTPYVSSSTAAANYTSSFSSSSEVCTECTETESTST 412

Qy       427 VVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFAT 486
Db       413 STPYV--------TSSSWSSSEVCTECTE-----TESTSYVTPYVSSSTAAANYTSSFSS 459

Qy       487 TTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTV 546
Db       460 SSEVCTECTETES---------TSTSTPYVTSSSSSSEVCTECTETESTSYVTPYVSSS 510

Qy       547 TTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTD 606
Db       511 TAAANYTSSFSSSSEVCTECTETES---------TSTSTPYVTSSSWSSSEVCTECTETE 561

Qy       607 TVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYAT 666
Db       562 STSYVTPYVSSSTAAANYTSSFSSSSEVCTECTETES---------TSTSTPYATSSTGT 612

Qy       667 TTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTV 726
Db       613 ATSFTASTSNTMTSLVQ------TDTTVSFSLSSTVSEHTNAP-----TSSVESNASTFI 661
```

```
Qy       727 TTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVT-----TTEYWSQSFATTTTVTAP 781
             :: :   :|| |::  :  |          | | |||     :|   |:|  :: :||
Db       662 SSNKGSVKSYVTSSIHSITP--------MYPSNQTVTSSSVVSTPITSESSESSASVTIL 713

Qy       782 PGGTDTVIIYESMSSSKISTSSNDITSIIPSF---SRPHYVNSTTSDLSTFESSSMNTPT 838
             |  | |     |   :|: : |:  |::| |:   |:   | |:|| :|   | |: :
Db       714 P-STITSEFKPSTMKTKVVSISSSPTNLITSYDTTSKDSTVGSSTSSVSLISSISLPSSY 772

Qy       839 SISSD----GMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTST 894
             | ||:      ::||      | ::|::  |:  |  | | ::       |:|| | :|
Db       773 SASSEQIFHSSIVSSNGQALTSFSSTKVSSSESSESHRTSPTT-------SSESGIKSSG 825

Qy       895 VPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSS 954
             |   || : | |  :  ::|:| :|  |   ||  : ::|:     | |: : | |:
Db       826 VEIESTSTSSFSFHE--TSTASTSVQISSQFVTPSSPISTVA----PRSTGLNSQTE--- 876

Qy       955 IPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTD------- 1007
                 | : || | :|  :: |    | :|| :   ||:  |       ||
Db       877 ----STNSSKETMSSE-NSASVMPSSSATSPKTGKVTSDET-------SSGFSRDRTTVY 924

Qy      1008 -LTSEPTNTREQPTTLSTT-------SNSITEDITTSQPTGDNGDNTSSTNPVPTVATST 1059
              :|||  :| || | :: :       ||:::  : ::  |  ||  |  |   |  ||
Db       925 RMTSETPSTNEQTTLITVSSCESNSCSNTVSSAVVSTATTTINGITTEYTTWCPLSATE- 983

Qy      1060 LASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAP---SPAVSSGTD 1116
             | : |: :::              | : | | |  |:   | | :|   | | ::  |
Db       984 LTTVSKLESE---------------EKTTLITVTSCESGVCSETASPAIVSTATATVND 1027

Qy      1117 VTTEPTDTREQPTT-LSTTSKTNSEL--------VATTQATNENGGKSPSTDLTSSLT-T 1166
             | |  :    | |  |: :|   :           | |:: : |    |::  ||  | |
Db      1028 VVTVYSTWSPQATNKLAVSSDIENSASKASFVSEAAETKSISRNNNFVPTSGTTSIETHT 1087

Qy      1167 GTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSI-VSNTPQTTLSQQVTSSS 1225
              |::: | ||: |           |||   | | ||||   | ||  |: | :  :  |
Db      1088 TTTSNASENSDNV----------SASEAVS-SKSVTNPVLISVSQQPRGTPASSMIGS- 1134

Qy      1226 PSTNTFIASTYDGSGSIIQHSTWLYGLITL--LSLFI 1260
               || :   |:| |   |  |      |:|   :|:||
Db      1135 -STASLEMSSYLG---IANH------LLTNSGISIFI 1161
```

Search completed: June 11, 2003, 17:12:10
Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 17:12:16 ; Search time 298 Seconds
(without alignments)
2726.053 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQQFTLLFLYLSIASAKTI.....SIIQHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:*
1: /cgn2_6/ptodata/1/paa/PCTUS_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US080_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/1/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/1/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/1/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/1/paa/US100_COMB.pep:*
25: /cgn2_6/ptodata/1/paa/US101_COMB.pep:*
26: /cgn2_6/ptodata/1/paa/US102_COMB.pep:*
27: /cgn2_6/ptodata/1/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6495	100.0	1260	21	US-09-715-876-8	Sequence 8, Appli
2	6465	99.5	1260	27	US-60-385-568-211	Sequence 211, App
3	4204.5	64.7	1119	27	US-60-385-568-213	Sequence 213, App
4	4190.5	64.5	1191	16	US-09-248-796-16243	Sequence 16243, A
5	4190.5	64.5	1191	27	US-60-096-409-16243	Sequence 16243, A
6	4105	63.2	828	25	US-10-179-131-7791	Sequence 7791, Ap
7	4078.5	62.8	1047	27	US-60-385-568-218	Sequence 218, App
8	3951	60.8	1586	27	US-60-385-568-295	Sequence 295, App
9	3777.5	58.2	885	25	US-10-179-131-8318	Sequence 8318, Ap
10	3777.5	58.2	886	27	US-60-385-568-297	Sequence 297, App
11	3765.5	58.0	1756	27	US-60-385-568-322	Sequence 322, App
12	3739	57.6	1182	25	US-10-179-131-7998	Sequence 7998, Ap
13	3736.5	57.5	1581	25	US-10-179-131-10126	Sequence 10126, A
14	3716.5	57.2	1523	27	US-60-385-568-214	Sequence 214, App
15	3620	55.7	1593	27	US-60-385-568-323	Sequence 323, App
16	3586	55.2	1232	25	US-10-179-131-7938	Sequence 7938, Ap
17	3375	52.0	657	25	US-10-179-131-6795	Sequence 6795, Ap
18	3220	49.6	1060	16	US-09-248-796-14123	Sequence 14123, A
19	3220	49.6	1060	27	US-60-096-409-14123	Sequence 14123, A
20	3194	49.2	1419	27	US-60-385-568-215	Sequence 215, App
21	3067.5	47.2	952	27	US-60-385-568-319	Sequence 319, App
22	3067	47.2	894	27	US-60-385-568-287	Sequence 287, App
23	3015.5	46.4	819	25	US-10-179-131-5983	Sequence 5983, Ap
24	2937	45.2	1347	27	US-60-385-568-361	Sequence 361, App
25	2791.5	43.0	1443	27	US-60-385-568-216	Sequence 216, App
26	2779	42.8	780	16	US-09-248-796-16702	Sequence 16702, A
27	2779	42.8	780	27	US-60-096-409-16702	Sequence 16702, A
28	2623	40.4	775	25	US-10-179-131-8127	Sequence 8127, Ap
29	2582	39.8	1029	27	US-60-385-568-293	Sequence 293, App
30	2576.5	39.7	2297	27	US-60-385-568-217	Sequence 217, App
31	2560	39.4	1189	25	US-10-179-131-6661	Sequence 6661, Ap
32	2548.5	39.2	2000	27	US-60-385-568-409	Sequence 409, App
33	2544	39.2	1366	27	US-60-385-568-411	Sequence 411, App
34	2544	39.2	1384	25	US-10-179-131-9947	Sequence 9947, Ap
35	2320	35.7	909	25	US-10-179-131-6405	Sequence 6405, Ap
36	2297	35.4	646	16	US-09-248-796-16701	Sequence 16701, A
37	2297	35.4	646	27	US-60-096-409-16701	Sequence 16701, A
38	2273	35.0	812	16	US-09-248-796-16244	Sequence 16244, A
39	2273	35.0	812	27	US-60-096-409-16244	Sequence 16244, A
40	2271	35.0	467	27	US-60-385-568-288	Sequence 288, App
41	2243.5	34.5	895	25	US-10-179-131-5330	Sequence 5330, Ap
42	1994	30.7	511	16	US-09-248-796-14125	Sequence 14125, A
43	1994	30.7	511	27	US-60-096-409-14125	Sequence 14125, A
44	1986.5	30.6	468	27	US-60-385-568-212	Sequence 212, App
45	1973	30.4	451	25	US-10-179-131-5151	Sequence 5151, Ap

ALIGNMENTS

RESULT 1
US-09-715-876-8
; Sequence 8, Application US/09715876
; GENERAL INFORMATION:
; APPLICANT: Edwards Jr., John E., et al.
; TITLE OF INVENTION: Pharmaceutical Compositions and Methods to Vaccinate Against C
; TITLE OF INVENTION: Candidiasis.
; FILE REFERENCE: 259/064
; CURRENT APPLICATION NUMBER: US/09/715,876
; CURRENT FILING DATE: 2000-11-18
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 1260
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-715-876-8

Query Match 100.0%; Score 6495; DB 21; Length 1260;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
```

```
Qy       121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db       121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180

Qy       181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
Db       181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240

Qy       241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db       241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300

Qy       301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
Db       301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

Qy       361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
Db       361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy       421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
Db       421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480

Qy       481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
Db       481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540

Qy       541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
Db       541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600

Qy       601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
Db       601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660

Qy       661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
Db       661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720

Qy       721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780
Db       721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780

Qy       781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840
Db       781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840

Qy       841 SSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900
Db       841 SSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900

Qy       901 MSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE 960
Db       901 MSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE 960

Qy       961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT 1020
Db       961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT 1020

Qy      1021 TLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080
Db      1021 TLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080

Qy      1081 KPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE 1140
Db      1081 KPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE 1140

Qy      1141 LVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHST 1200
Db      1141 LVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHST 1200

Qy      1201 SVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db      1201 SVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
```

RESULT 2
US-60-385-568-211
; Sequence 211, Application US/60385568
; GENERAL INFORMATION:
; APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Giandong Zeng
; TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and Their Use
; FILE REFERENCE: 032796-134
; CURRENT APPLICATION NUMBER: US/60/385,568
; CURRENT FILING DATE: 2002-06-05
; PRIOR FILING DATE: June 5, 2002
; NUMBER OF SEQ ID NOS: 418
; SEQ ID NO 211
; LENGTH: 1260
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-385-568-211

Query Match 99.5%; Score 6465; DB 27; Length 1260;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1255; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy         1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
Db         1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60

Qy        61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
Db        61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120

Qy       121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db       121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180

Qy       181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
Db       181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240

Qy       241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db       241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300

Qy       301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
Db       301 SRSQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

Qy       361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
Db       361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy       421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
Db       421 TDSIDTVVVQVPSPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480

Qy       481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
Db       481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540

Qy       541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
Db       541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600

Qy       601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
Db       601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660

Qy       661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
Db       661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
```

```
Qy       721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780
Db       721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780

Qy       781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840
Db       781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840

Qy       841 SSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900
Db       841 SSDGMLLSSTTLVTESETTTESICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900

Qy       901 MSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE 960
Db       901 MSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE 960

Qy       961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT 1020
Db       961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT 1020

Qy      1021 TLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080
Db      1021 TLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080

Qy      1081 KPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE 1140
Db      1081 KPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE 1140

Qy      1141 LVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHST 1200
Db      1141 SVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHST 1200

Qy      1201 SVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db      1201 SVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
```

RESULT 3
US-60-385-568-213
; Sequence 213, Application US/60385568
; GENERAL INFORMATION:
; APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Giandong Zeng
; TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and Their Use
; FILE REFERENCE: 032796-134
; CURRENT APPLICATION NUMBER: US/60/385,568
; CURRENT FILING DATE: 2002-06-05
; PRIOR FILING DATE: June 5, 2002
; NUMBER OF SEQ ID NOS: 418
; SEQ ID NO 213
; LENGTH: 1119
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-385-568-213

Query Match 64.7%; Score 4204.5; DB 27; Length 1119;
Best Local Similarity 64.9%; Pred. No. 1.4e-297;
Matches 844; Conservative 106; Mismatches 128; Indels 223; Gaps 18;

```
Qy         1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
Db         1 MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYNYKGPGTPTWNAVLGWSLDGTS 60

Qy        61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
Db        61 ASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGEEFMTFSTLTCTVSNTLTP 120

Qy       121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db       121 SIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTFNDGGKKISINVDFERSNVDP 180

Qy       181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
Db       181 KGYLTDSRVIPSLNKVSTLFVAPQCANGYTSGTMGFANTYGDVQIDCSNIHVGITKGLND 240

Qy       241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db       241 WNYPVSSESFSYTKTCSSNGIFITYKNVPAGYRPFVDAYISATDVNSYTLSYANEYTCAG 300

Qy       301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
Db       301 GYWQRAPFTLRWTGYRNSDAGSNGIVIVATTRTVTDSTTAVTTLPFDPNRDKTKTIEILK 360

Qy       361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
Db       361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDIPYHTTTTVTSKWTGTITSTTTHTNP 420

Qy       421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
Db       421 TDSIDTVIVQVPSPNPTVTTTEYWSQSFATTTTITGPPGNTDTVLIREPPNHTVTTTEYW 480

Qy       481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
Db       481 SESYTTTSTFTAPPGGTDSVIIKEPPNPTVTTTEYWSESYTTTSTFTAPPGGTDSVIIKE 540

Qy       541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
Db       541 PPNHTVTTTEYWSQSYTTTTTVTAPPGGTDTVLVREPPNHTVTTTEYWSQSYTTTTTVIA 600

Qy       601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
Db       601 PPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYW 660

Qy       661 SQSYATTTTITAPPGETDTVL----------------------------------------IRE 684
Db       661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDTVIIRE 720

Qy       685 PPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTA 744
Db       721 PPNHTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTIIA 780

Qy       745 PPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSN 804
Db       781 PPGETDTVLIREPPNPTVTTTEYWSQSYTTATTVTAPPGGTDTVIIYDTMSSSEIS---- 836

Qy       805 DITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELIC 864
Db       837 -------SFSRPHYTNHTT----------------------LWSTTWVIETKTITETSC 866

Qy       865 SDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGV 924
Db       867 EGDKGCSWVSVSTRIVTIPNNIETPMVTNTVDSTTTESTS-QSPSGI-------FSESGV 918

Qy       925 SVTTETS-VTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQST 983
Db       919 SVETESSTVTTAQTN----------------PSVPTTESEVVFTTKGNNENGPYESPST 961

Qy       984 STEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNG 1043
Db       962 N----------------VKSSMDENSEFT---------------------------- 974

Qy      1044 DNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIE---ATT 1100
Db       975 -------------TSTAAS----------------------------TSTDIENETIAT 992

Qy      1101 TSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSELVATTQATNEN-GGKSPSTD 1159
Db       993 TGSVEASSPIISSSADET----------TTVTTTAESTSVI---EQPTNNNGGGKAPSA- 1038

Qy      1160 LTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQ 1219
Db       1039 -----TSSPSTTTTANNDSVITG--------TTSTNQSQSQSQYN-----SDTQQTTLSQ 1080

Qy      1220 QVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db       1081 QMTSSLVSLH--MLTTFDGSGSVIQHSTWLCGLITLLSLFI 1119
```

```
RESULT 4
US-09-248-796-16243
; Sequence 16243, Application US/09248796
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
;  TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796
;  CURRENT FILING DATE: 1999-02-12
;  NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 16243
;   LENGTH: 1191
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796-16243

  Query Match             64.5%;  Score 4190.5;  DB 16;  Length 1191;
  Best Local Similarity   68.1%;  Pred. No. 1.6e-296;
  Matches  863;  Conservative   89;  Mismatches  195;  Indels  121;  Gaps   19;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 62

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db         63 ANPGDTFTLNMPCVFRYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 122

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              |||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db        123 SIKAFGTVTLPIAFNVGGTGSSTDVEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 182

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 SGYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 242

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 302

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 362

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 422

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
              || |||||||||||||:||||||||:||||||||||||||:|||||||| ||||||||
Db        423 TDYIDTVVVQVPLPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYW 482

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
              |||:||::|||||||||||:|||||||||||||||||||:|||||||||||||||||||
Db        483 SQSYATSSTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIRE 542

Qy        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
              |||||||||||||||:||||:|||||  ||:|:|||||||||||||||||  ||:|:||
Db        543 PPNPTVTTTEYWSQSFATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYVTTSTITA 602

Qy        601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
              |||||||||||||||:|||||||||||:|||||||  || ||||||||||||||||||||
Db        603 PPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYW 662

Qy        661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
              |||||||||:| ||| ||||:||||||  ||||||||||||||||||||  ||||:|||
Db        663 SQSYATTTTVTGPPGGTDTVIIREPPNQTVTTTEYWSQSYATTTTVTAPPGGTDTVIIRE 722

Qy        721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780
              ||| ||||||||||||||||| ||||||||||||||||||||||:||||||||
Db        723 PPNPTVTTTEYWSQSYATTTTVTGPPGGTDTVIIREPPNPTVTTTEYWSQSYATTTTVTA 782

Qy        781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840
              ||||| |||| |                     |     ||::   : | |  | |::
Db        783 PPGGTATVIIRE----------------------PPNPTVTTTE---YWSQSYATTTTV 816

Qy        841 SS--DGMLLSSTTLVTE----SETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTST 894
              :    |   : | :: |    : ||||           | |    |          | |
Db        817 TGPPGG---TDTVIIREPPNPTVTTTEY----------WSQSYATTT----------TVT 853

Qy        895 VPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSS 954
              |   | :  :       : |:::  |:|  : || |          |   : | : :
Db        854 APPGGTATVIIREPPNYTVTTTEYWSQSYATTTTVT---------GPPGGTDTVIIREPP 904

Qy        955 IPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTN 1014
              |:|: :|                : ||| :|         :| |  |   : | :  || |
Db        905 SPTVTTTE---------------YWSQSYAT--------TTTVTAPPGGTATVIIWEPPN 941

Qy       1015 TREQPTTLSTT--SNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGS 1072
                  ||  :|   | |      | : | |          | ||| |:        :  | |
Db        942 ----PTVTTTEYWSQSYATTTTVTGPPGGTDTVIIREPPSPTVTTT--------EYWSQS 989

Qy       1073 HESASTSLKPSMGENSGLTTST----EIEATTTSPTEAPSPAVSSGTDVTTEP--TDT-- 1124
              : : :|   |  |      |:|    |    | : ||  | : :: | ||  |  |||
Db        990 YATTTTVTAPPGG------TATVIIREPPNYTVTTTEYWSQSYATTTTVTGPPGGTDTVI 1043

Qy       1125 -REQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVT--- 1180
               || |    ||::  |:  |||       | : |  :    :|   :| |  | |
Db       1044 IREPPNPTVTTTEYWSQSYATTLTITAPPGGTNSVIIRVHSSTNDESSESTFSTLSVPSF 1103

Qy       1181 SGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSG 1240
              |||::   |:: |     :::||:  | :|       |  |||:: :: | :  :  :|
Db       1104 SGSIS--VVSTISRPHYVNSTVTHLPSSLSKPVDIPSSDAVTSTNDNSLTSLTGSENGKT 1161

Qy       1241 SIIQHSTW 1248
              |:   :|:
Db       1162 SVAISTTF 1169


RESULT 5
US-60-096-409-16243
; Sequence 16243, Application US/60096409A
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
;  TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-016P
;  CURRENT APPLICATION NUMBER: US/60/096,409A
;  CURRENT FILING DATE: 1998-08-13
;  NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 16243
;   LENGTH: 1191
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-60-096-409-16243

  Query Match             64.5%;  Score 4190.5;  DB 27;  Length 1191;
  Best Local Similarity   68.1%;  Pred. No. 1.6e-296;
  Matches  863;  Conservative   89;  Mismatches  195;  Indels  121;  Gaps   19;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 62

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              |||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db         63 ANPGDTFTLNMPCVFRYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 122

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              |||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db        123 SIKAFGTVTLPIAFNVGGTGSSTDVEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 182

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
```

```
Db        183 SGYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 242

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 302

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 362

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 422

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
              || ||||||||||||||:||||||||:||||||||||||||:|||||||| ||||||||
Db        423 TDYIDTVVVQVPLPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYW 482

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
              |||:||::||||||||||:|||||||||||||||||:|||||||||||||||||||||
Db        483 SQSYATSSTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIRE 542

Qy        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
              ||||||||||||||:|||||:||||| ||:|:|||||||||||||||| ||:|:||
Db        543 PPNPTVTTTEYWSQSFATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYVTTSTITA 602

Qy        601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
              ||||||||||||||:|||||||||:|||||||| || ||||||||||||||||||||||
Db        603 PPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYW 662

Qy        661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
              |||||||||:| ||| |||||:|||||| ||||||||||||||||||||| ||||:|||
Db        663 SQSYATTTTVTGPPGGTDTVIIREPPNQTVTTTEYWSQSYATTTTVTAPPGGTDTVIIRE 722

Qy        721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780
              ||| |||||||||||||||||| ||||||||||||||||||||||||:|||||||||
Db        723 PPNPTVTTTEYWSQSYATTTTVTGPPGGTDTVIIREPPNPTVTTTEYWSQSYATTTTVTA 782

Qy        781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840
              ||||| |||| |                        |     ||::   : | |  | |::
Db        783 PPGGTATVIIRE----------------------PPNPTVTTTE---YWSQSYATTTTV 816

Qy        841 SS--DGMLLSSTTLVTE----SETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTST 894
              :    |   : | :: |    : ||||            | |    |          | |
Db        817 TGPPGG---TDTVIIREPPNPTVTTTEY----------WSQSYATTT----------TVT 853

Qy        895 VPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSS 954
              |   | :  :      : |:::  |:|  : || |        |   : | : :
Db        854 APPGGTATVIIREPPNYTVTTTEYWSQSYATTTTVT---------GPPGGTDTVIIREPP 904

Qy        955 IPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTN 1014
              |:|: :|              : ||| :|        :| |  |   : | :  || |
Db        905 SPTVTTTE--------------YWSQSYAT--------TTTVTAPPGGTATVIIWEPPN 941

Qy       1015 TREQPTTLSTT--SNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGS 1072
                ||  :|   | |     | : | |          | ||| |:        :  | |
Db        942 ----PTVTTTEYWSQSYATTTTVTGPPGGTDTVIIREPPSPTVTTT--------EYWSQS 989

Qy       1073 HESASTSLKPSMGENSGLTTST----EIEATTTSPTEAPSPAVSSGTDVTTEP--TDT-- 1124
              : : :|   |  |      |:|    |    | : ||  | : :: | ||  |  |||
Db        990 YATTTTVTAPPGG------TATVIIREPPNYTVTTTEYWSQSYATTTTVTGPPGGTDTVI 1043

Qy       1125 -REQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVT--- 1180
               || |    ||::  |:  |||       | : |  :    :|   :| |  | |
Db       1044 IREPPNPTVTTTEYWSQSYATTLTITAPPGGTNSVIIRVHSSTNDESSESTFSTLSVPSF 1103

Qy       1181 SGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSG 1240
              |||::   |:: |     :::||:  | :|       |  |||:: :: | :  : :|
Db       1104 SGSIS--VVSTISRPHYVNSTVTHLPSSLSKPVDIPSSDAVTSTNDNSLTSLTGSENGKT 1161

Qy       1241 SIIQHSTW 1248
              |:   :|:
```

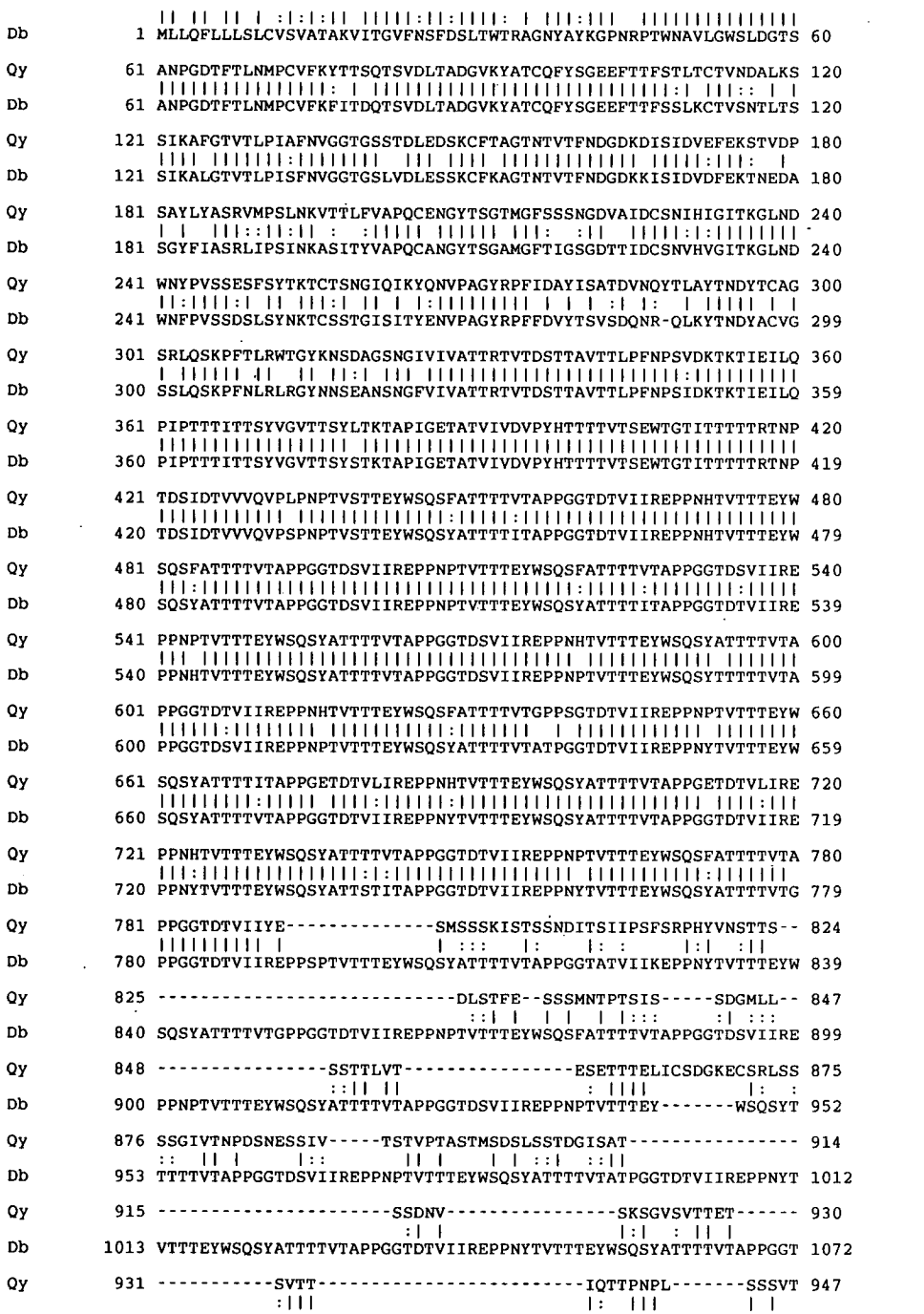
```
Db       1162 SVAISTTF 1169
```

RESULT 6
US-10-179-131-7791
; Sequence 7791, Application US/10179131
; GENERAL INFORMATION:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHIMER JR., GEORGE H.
; APPLICANT: KESSLER, MARCO
; APPLICANT: NOLLING, JORK
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: 2976-4031
; CURRENT APPLICATION NUMBER: US/10/179,131
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 10194
; SEQ ID NO 7791
; LENGTH: 828
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-179-131-7791

Query Match 63.2%; Score 4105; DB 25; Length 828;
Best Local Similarity 97.1%; Pred. No. 1.7e-290;
Matches 803; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

```
Qy        434 PNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAP 493
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db          2 PNPTVSTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAP 61

Qy        494 PGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWS 553
              || ||:|||||||| ||||||||||:|||||||||||||:|:|||||| |||||||||
Db         62 PGETDTVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVLIREPPNHTVTTTEYWS 121

Qy        554 QSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREP 613
              ||||||||||||||||:|||||||:|||||||||||||||||:||||| |||||||||
Db        122 QSYATTTTVTAPPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTITAPPGETDTVIIREP 181

Qy        614 PNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAP 673
              ||||||||||||||:|||||| || |||||:|||||| ||||||||||||||||:|||
Db        182 PNHTVTTTEYWSQSYATTTTVTAPPGGTDTVLIREPPNHTVTTTEYWSQSYATTTTVTAP 241

Qy        674 PGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWS 733
              || |||||||||||||||||||||||||||||||||| ||||:||||||||||||||||
Db        242 PGGTDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWS 301

Qy        734 QSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYES 793
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        302 QSYATTTTVTAPPGGTDTVIIREPPNYTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYES 361

Qy        794 MSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLV 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 MSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLV 421

Qy        854 TESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISA 913
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TESETTTESICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISA 481

Qy        914 TSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDN 973
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 TSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDN 541

Qy        974 QSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDI 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 QSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDI 601

Qy       1034 TTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTS 1093
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 TTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTS 661
```

```
Qy       1094 TEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSELVATTQATNENGG 1153
Db        662 TEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSESVATTQATNENGG 721

Qy       1154 KSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTP 1213
Db        722 KSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTP 781

Qy       1214 QTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db        782 QTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 828


RESULT 7
US-60-385-568-218
; Sequence 218, Application US/60385568
; GENERAL INFORMATION:
;  APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Giandong Zeng
;  TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and Their Use
;  FILE REFERENCE: 032796-134
;  CURRENT APPLICATION NUMBER: US/60/385,568
;  CURRENT FILING DATE:  2002-06-05
;  PRIOR FILING DATE: June 5, 2002
;  NUMBER OF SEQ ID NOS: 418
; SEQ ID NO 218
;   LENGTH: 1047
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-60-385-568-218

  Query Match             62.8%;  Score 4078.5;  DB 27;  Length 1047;
  Best Local Similarity   65.0%;  Pred. No. 2.1e-288;
  Matches  822;  Conservative  100;  Mismatches  120;  Indels  223;  Gaps   18;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
Db          1 MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYHYKGPGTPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
Db         61 ASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGEEFMTFSTLTCTVSNTLTP 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db        121 SIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTFNDGGKKISINVDFERSNVDP 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
Db        181 KGYLTDSRVIPSLNKVSTLFVAPQCANGYTSGTMGFANTYGDVQIDCSNIHVGITKGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db        241 WNYPVSSESFSYTKTCSSNGIFITYKNVPAGYRPFVDAYISATDVNSYTLSYANEYTCAG 300

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
Db        301 GYWQRAPFTLRWTGYRNSDAGSNGIVIVATTRTVTDSTTAVTTLPFDPNRDKTKTIEILK 360

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
Db        361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDIPYHTTTTVTSKWTGTITSTTTHTNP 420

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
Db        421 TDSIDTVIVQVPSPNPTVTTTEYWSQSFATTTTITGPPGNTDTVLIREPPNHTVTTTEYW 480

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
Db        481 SESYTTTSTFTAPPGGTDSVIIKEPPNPTVTTTEYWSESYTTTSTFTAPPGGTDSVIIKE 540

Qy        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
Db        541 PPNPTVTTTEYWSQSYTTTTTVTAPPGGTDTVLVREPPNHTVTTTEYWSQSYTTTTTVIA 600

Qy        601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
Db        601 PPGGTDTVIIREPPNH-----------------------------------TVTTTEYW 624

Qy        661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
Db        625 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTITAPPGETDTVLIRE 684

Qy        721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780
Db        685 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVLIREPPNHTVTTTEYWSQSYATTTTVTA 744

Qy        781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840
Db        745 PPGGTDTVIIYDTMSSSEIS-----------SFSRPHYTNHTT----------------- 776

Qy        841 SSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900
Db        777 ------LWSTTWVIETKTITETSCEGDKGCSWVSVSTRIVTIPNNIETPMVTNTVDSTTT 830

Qy        901 MSDSLSSTDGISATSSDNVSKSGVSVTTETS-VTTIQTTPNPLSSSVTSLTQLSSIPSVS 959
Db        831 ESTS-QSPSGI-------FSESGVSVETESSTVTTAQTN----------------PSVP 865

Qy        960 ESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQP 1019
Db        866 TTESEVEFTTKGNNGNGPYESPSTH----------------VKSSMDENSEFT------ 902

Qy       1020 TTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTS 1079
Db        903 ------------------------------------TSTAAS----------------- 908

Qy       1080 LKPSMGENSGLTTSTEIE---ATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSK 1136
Db        909 -----------TSTDIENATIATTGSVEASSPIISSSADET----------TTITTTAE 946

Qy       1137 TNSELVATTQATNEN-GGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASND 1195
Db        947 STSVI---EQPTNNNGGGKAPSA------TSSPSTTTTANNDSVITG--------TTSTN 989

Qy       1196 QSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITL 1255
Db        990 QSQSQSQSN-----SDTQQTTLSQQMTSSLVSLH--MLTTFDGSGSVIQHSTWLCGLITL 1042

Qy       1256 LSLFI 1260
Db       1043 LSLFI 1047


RESULT 8
US-60-385-568-295
; Sequence 295, Application US/60385568
; GENERAL INFORMATION:
;  APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Giandong Zeng
;  TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and Their Use
;  FILE REFERENCE: 032796-134
;  CURRENT APPLICATION NUMBER: US/60/385,568
;  CURRENT FILING DATE:  2002-06-05
;  PRIOR FILING DATE: June 5, 2002
;  NUMBER OF SEQ ID NOS: 418
; SEQ ID NO 295
;   LENGTH: 1586
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-60-385-568-295

  Query Match             60.8%;  Score 3951;  DB 27;  Length 1586;
  Best Local Similarity   53.4%;  Pred. No. 8.3e-279;
  Matches  859;  Conservative  140;  Mismatches  238;  Indels  372;  Gaps   34;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
```



```
Db          1 MLLQFLLLSLCVSVATAKVITGVFNSFDSLTWTRAGNYAYKGPNRPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
Db         61 ANPGDTFTLNMPCVFKFITDQTSVDLTADGVKYATCQFYSGEEFTTFSSLKCTVSNTLTS 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db        121 SIKALGTVTLPISFNVGGTGSLVDLESSKCFKAGTNTVTFNDGDKKISIDVDFEKTNEDA 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
Db        181 SGYFIASRLIPSINKASITYVAPQCANGYTSGAMGFTIGSGDTTIDCSNVHVGITKGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db        241 WNFPVSSDSLSYNKTCSSTGISITYENVPAGYRPFFDVYTSVSDQNR-QLKYTNDYACVG 299

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
Db        300 SSLQSKPFNLRLRGYNNSEANSNGFVIVATTRTVTDSTTAVTTLPFNPSIDKTKTIEILQ 359

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
Db        360 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 419

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
Db        420 TDSIDTVVVQVPSPNPTVSTTEYWSQSYATTTTITAPPGGTDTVIIREPPNHTVTTTEYW 479

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
Db        480 SQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTITAPPGGTDTVIIRE 539

Qy        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
Db        540 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYTTTTTVTA 599

Qy        601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
Db        600 PPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTATPGGTDTVIIREPPNYTVTTTEYW 659

Qy        661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
Db        660 SQSYATTTTVTAPPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTAPPGGTDTVIIRE 719

Qy        721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780
Db        720 PPNYTVTTTEYWSQSYATTSTITAPPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTG 779

Qy        781 PPGGTDTVIIYE--------------SMSSSKISTSSNDITSIIPSFSRPHYVNSTTS-- 824
Db        780 PPGGTDTVIIREPPSPTVTTTEYWSQSYATTTTVTAPPGGTATVIIKEPPNYTVTTTEYW 839

Qy        825 ---------------------------DLSTFE--SSSMNTPTSIS-----SDGMLL-- 847
Db        840 SQSYATTTTVTGPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 899

Qy        848 ---------------SSTTLVT---------------ESETTTELICSDGKECSRLSS 875
Db        900 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEY-------WSQSYT 952

Qy        876 SSGIVTNPDSNESSIV-----TSTVPTASTMSDSLSSTDGISAT--------------- 914
Db        953 TTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTATPGGTDTVIIREPPNYT 1012

Qy        915 ---------------------SSDNV---------------SKSGVSVTTET------ 930
Db       1013 VTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTAPPGGT 1072

Qy        931 -----------SVTT------------------------IQTTPNPL-------SSSVT 947

Db       1073 DTVIIREPPNHTVTTTEYWSQSYTTTTTVIAPPGGTDSVIIREPPNPTVTTTEYWSQSYA 1132

Qy        948 SLTQLSSIPSVSE-----------------SESKVTFTSNGDNQSGT------------ 977
Db       1133 TTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHT 1192

Qy        978 -------HDSQSTSTEI-----------------EIVTT---------SSTKVLPPVVSS 1004
Db       1193 VTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTIIAPPGET 1252

Qy       1005 NTDLTSEPTNTREQPTTLSTT--SNSITEDITTSQPTG----------DNGDNTSSTNP 1051
Db       1253 DTVLIREPPN----PTVTTTEYWSQSYTTATTVTAPPGGTDTVIIYDTMSSSEISSFSRP 1308

Qy       1052 VPT-----------VATSTLASASEEDNKSGSHESASTSLK--PSMGENSGLTTSTEIEA 1098
Db       1309 HYTNHTTLWSTTWVIETKTITETSCEGDKGCSWVSVSTRIVTIPNNIE-TPMVTNTVDTT 1367

Qy       1099 TTTSPTEAPSPAVS-SGTDVTTEPTDTREQPTTLSTTSKTN-------SELVATTQATNE 1150
Db       1368 TTESTLQSPSGIFSESGVSVETE--------SSTFTTAQTNPSVPTTESEVVFTTKGNNG 1419

Qy       1151 NGG-KSPSTDLTSSL------TTGTSASTSA---NSELVTSGSVT--------------- 1185
Db       1420 NGPYESPSTNVKSSMDENSEFTTSTAASTSTDIENETIATTGSVEASSPIISSSADETTT 1479

Qy       1186 -------------------GGAVASASNDQSHSTSVT-NSNSIV-------------SN 1211
Db       1480 VTTTAESTSVIEQQTNNNGGGNAPSATSTSSPSTTTTANSDSVITSTTSTNQSQSQSNSD 1539

Qy       1212 TPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db       1540 TQQTTLSQQMTSSLVSLH--MLTTFDGSGSVIQHSTWLCGLITLLSLFI 1586
```

RESULT 9
US-10-179-131-8318
; Sequence 8318, Application US/10179131
; GENERAL INFORMATION:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHIMER JR., GEORGE H.
; APPLICANT: KESSLER, MARCO
; APPLICANT: NOLLING, JORK
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: 2976-4031
; CURRENT APPLICATION NUMBER: US/10/179,131
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 10194
; SEQ ID NO 8318
; LENGTH: 885
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-179-131-8318

Query Match 58.2%; Score 3777.5; DB 25; Length 885;
Best Local Similarity 79.0%; Pred. No. 1.6e-266;
Matches 707; Conservative 56; Mismatches 49; Indels 83; Gaps 3;

```
Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
Db          1 MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYNYKGPGTPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
Db         61 ASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGEEFMTFSTLTCTVSNTLTP 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db        121 SIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTFNDGGKKISINVDFERSNVDP 180
```

```
Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
Db        181 KGYLTDSRVIPSLNKVSTLFVAPQCANGYTSGTMGFANTYGDVQIDCSNIHVGITKGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db        241 WNYPVSSESFSYTKTCSSNGIFITYKNVPAGYRPFVDAYISATDVNSYTLSYANEYTCAG 300

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
Db        301 GYWQRAPFTLRWTGYRNSDAGSNGIVIVATTRTVTDSTTAVTTLPFDPNRDKTKTIEILK 360

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
Db        361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDIPYHTTTTVTSKWTGTITSTTTHTNP 420

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
Db        421 TDSIDTVIVQVPSPNPTVTTTEYWSQSFATTTTITGPPGNTDTVLIREPPNHTVTTTEYW 480

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVT------------- 527
Db        481 SESYTTTSTFTAPPGGTDSVIIKEPPNPTVTTTEYWSESYTTTTTVTAPPGGTDTVIIRE 540

Qy        528 ----------------------APPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTA 564
Db        541 PPNHTVTTTEYWSQSYTTTTTVIAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTITA 600

Qy        565 PPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 624
Db        601 PPGETDTVLIREPPNHTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYW 660

Qy        625 SQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIRE 684
Db        661 SQSYTTTTTVIAPPGGTDSVIIKEPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIRE 720

Qy        685 PPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTA 744
Db        721 PPNHTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSFATTTTVTA 780

Qy        745 PPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVT------------------------- 779
Db        781 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDTVLIREPPNPTVTTTEYW 840

Qy        780 -----------APPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTT 823
Db        841 SQPYTTTTTVIAPPGGTDTVIIYDTMSSSEIS-----------SFSRPHYTNHTT 884
```

RESULT 10
US-60-385-568-297
; Sequence 297, Application US/60385568
; GENERAL INFORMATION:
; APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Giandong Zeng
; TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and Their Use
; FILE REFERENCE: 032796-134
; CURRENT APPLICATION NUMBER: US/60/385,568
; CURRENT FILING DATE: 2002-06-05
; PRIOR FILING DATE: June 5, 2002
; NUMBER OF SEQ ID NOS: 418
; SEQ ID NO 297
; LENGTH: 886
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-385-568-297

Query Match 58.2%; Score 3777.5; DB 27; Length 886;
Best Local Similarity 79.0%; Pred. No. 1.6e-266;
Matches 707; Conservative 56; Mismatches 49; Indels 83; Gaps 3;

```
Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
Db          1 MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYNYKGPGTPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
Db         61 ASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGEEFMTFSTLTCTVSNTLTP 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db        121 SIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTFNDGGKKISINVDFERSNVDP 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
Db        181 KGYLTDSRVIPSLNKVSTLFVAPQCANGYTSGTMGFANTYGDVQIDCSNIHVGITKGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db        241 WNYPVSSESFSYTKTCSSNGIFITYKNVPAGYRPFVDAYISATDVNSYTLSYANEYTCAG 300

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
Db        301 GYWQRAPFTLRWTGYRNSDAGSNGIVIVATTRTVTDSTTAVTTLPFDPNRDKTKTIEILK 360

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
Db        361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDIPYHTTTTVTSKWTGTITSTTTHTNP 420

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
Db        421 TDSIDTVIVQVPSPNPTVTTTEYWSQSFATTTTITGPPGNTDTVLIREPPNHTVTTTEYW 480

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVT------------- 527
Db        481 SESYTTTSTFTAPPGGTDSVIIKEPPNPTVTTTEYWSESYTTTTTVTAPPGGTDTVIIRE 540

Qy        528 ----------------------APPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTA 564
Db        541 PPNHTVTTTEYWSQSYTTTTTVIAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTITA 600

Qy        565 PPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 624
Db        601 PPGETDTVLIREPPNHTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYW 660

Qy        625 SQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIRE 684
Db        661 SQSYTTTTTVIAPPGGTDSVIIKEPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIRE 720

Qy        685 PPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTA 744
Db        721 PPNHTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSFATTTTVTA 780

Qy        745 PPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVT------------------------- 779
Db        781 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDTVLIREPPNPTVTTTEYW 840

Qy        780 -----------APPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTT 823
Db        841 SQPYTTTTTVIAPPGGTDTVIIYDTMSSSEIS-----------SFSRPHYTNHTT 884
```

RESULT 11
US-60-385-568-322
; Sequence 322, Application US/60385568
; GENERAL INFORMATION:
; APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Giandong Zeng
; TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and Their Use
; FILE REFERENCE: 032796-134
; CURRENT APPLICATION NUMBER: US/60/385,568
; CURRENT FILING DATE: 2002-06-05
; PRIOR FILING DATE: June 5, 2002
; NUMBER OF SEQ ID NOS: 418
; SEQ ID NO 322
; LENGTH: 1756
; TYPE: PRT

```
;   ORGANISM: Candida albicans
US-60-385-568-322

  Query Match            58.0%;  Score 3765.5;  DB 27;  Length 1756;
  Best Local Similarity  58.4%;  Pred. No. 3.5e-265;
  Matches  789;  Conservative  136;  Mismatches  260;  Indels  165;  Gaps   29;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              || || || | :|:|:|| ||||||:||:||||: | |||:|||  ||||||||||||||
Db          1 MLLQFLLLSLCVSVATAKVITGVFNSFDSLTWTRAGNYAYKGPNRPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              |||||||||||||||: | ||||||||:|||||||||||||||||||:| |||:: | |
Db         61 ANPGDTFTLNMPCVFKFITDQTSVDLTAEGVKYATCQFYSGEEFTTFSSLKCTVSNTLTS 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              |||| |||||||:||||||||| ||| |:|| |||||||||||| |||||:|||:  |
Db        121 SIKALGTVTLPISFNVGGTGSSVDLESSQCFKAGTNTVTFNDGDKKISIDVDFEKTNEDA 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              | |  |||::||:|||:  :||||| |||||| |||    ||  |||||:|:||||||||
Db        181 SGYFIASRLIPSINKVSITYVAPQCANGYTSGAMGFIVLTGDTTIDCSNVHVGITKGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||:||||:| || |||:| || | |:|||||||| | | | :  |:  | ||||| | |
Db        241 WNFPVSSDSLSYNKTCSSTGISITYENVPAGYRPFFDVYTSVSGQNR-QLRYTNDYACVG 299

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
              | |||||| ||  || ||:| ||| ||||||||||||||||||||||||||||||||||||
Db        300 SSLQSKPFNLRLRGYNNSEANSNGFVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 359

Qy        361 PIPTTTIITSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        360 PIPTTTIITSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 419

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
              |||||||||||| |||||:||||||||:||||||||||||||:|||||||| ||||||||
Db        420 TDSIDTVVVQVPSPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYW 479

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
              |||:||:::|||||||||:|||||||||||||||||||:|||||||||||||||||||||
Db        480 SQSYATSSTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIRE 539

Qy        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
              |||||||||||||||:|||||:||||| ||:|:|||||||||||||||||| ||:|:||
Db        540 PPNPTVTTTEYWSQSFATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYVTTSTITA 599

Qy        601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
              |||||||||||||||:|||||||||||:||||||| || ||||||||||||||||||||
Db        600 PPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYW 659

Qy        661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
              |||||||||:| ||| ||||:|||||| |||||||||||||||||||||| ||||:|||
Db        660 SQSYATTTTVTGPPGGTDTVIIREPPNQTVTTTEYWSQSYATTTTVTAPPGGTDTVIIRE 719

Qy        721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780
              ||| |||||||||||||||||| ||||||||||||||||||||||||:||||||||
Db        720 PPNPTVTTTEYWSQSYATTTTVTGPPGGTDTVIIREPPNPTVTTTEYWSQSYATTTTVTA 779

Qy        781 PPGGTDTVIIYE----SMSSSKISTSSNDITSIIPS---------FSRPHYVNSTTSDLS 827
              ||||| |||| |    ::::::  : |   |: :              |     ||::
Db        780 PPGGTATVIIREPPNPTVTTTEYWSQSYATTTTVTGPPGGTDTVIIREPPNPTVTTTE-- 837

Qy        828 TFESSSMNTPTSISS--DGMLLSSTTLVTE----SETTTELICSDGKECSRLSSSSGIVT 881
              : | |  | |::::   |   ::| :: |    : ||||          |:  :::  ||
Db        838 -YWSQSYATTTTVTAPPGG---TATVIIREPPNYTVTTTEY-------WSQSYATTTTVT 886

Qy        882 NPDSNESSIV-----TSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQ 936
              |     :::     : || |    | | ::|  ::|                |:   |:
Db        887 GPPGGTDTVIIREPPSPTVTTTEYWSQSYATTTTVTAPPGG-----------TATVIIR 934

Qy        937 TTPNPL-------SSSVTSLTQLSSIPSVSES-------ESKVTFTSNGDNQSGTHDSQS 982
                |||        | |  : | ::  |  :::          || |         : |||

Db        935 EPPNPTVTTTEYWSQSYATTTTVTGPPGGTDTVIIREPPNPTVTTTE--------YWSQS 986

Qy        983 TSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTT---SNSITEDITTSQPT 1039
               :|         :| |  |    : | :  || |      |::||   | |      | : |
Db        987 YAT--------TTVTAPPGGTATVIIREPPN-----YTVTTTEYWSQSYATTTTVTGPP 1033

Qy       1040 GDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMG---------ENSGL 1090
              |          | ||| |:        :  | |: : :|   |  |           | :
Db       1034 GGTDTVIIREPPSPTVTTT--------EYWSQSYATTTTVTAPPGGTATVIIWEPPNPTV 1085

Qy       1091 TTS---TEIEATTTSPT------------EAPSPAV----------SSGTDVTTEPTDT- 1124
              ||:   :: . ||||: |             | ||| |          :: | ||  |  |
Db       1086 TTTEYWSQSYATTTTVTGPPGGTDTVIIREPPSPTVTTTEYWSQSYATTTTVTAPPGGTA 1145

Qy       1125 ----REQPT-TLSTTSKTNSELVATTQATNENGG-------KSPSTDLTSSLTTGTSAST 1172
                  || |  |::||   :    ||  |  ||       : |:  :|  ||  : :
Db       1146 TVIIREPPNYTVTTTEYWSQSYATTTTVTGPPGGTDTVIIREPPNPTVT---TTEYWSQS 1202

Qy       1173 SANSELVTS--GSVTGGAVASASN---------DQSHSTSVT------NSNSIVSNTP-- 1213
                | :  ||:  |      :    |          ||::|: |       ::|::   |
Db       1203 FATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPN 1262

Qy       1214 QTTLSQQVTSSSPSTNTFIASTYDGSGSII 1243
                |  : :  | | :| | : :   |: |:|
Db       1263 PTVTTTEYWSQSFATTTTVTAPPGGTDSVI 1292


RESULT 12
US-10-179-131-7998
; Sequence 7998, Application US/10179131
; GENERAL INFORMATION:
;  APPLICANT: HARE, ROBERTA S.
;  APPLICANT:  SHAW, KAREN J.
;  APPLICANT:  SHIMER JR., GEORGE H.
;  APPLICANT:  KESSLER, MARCO
;  APPLICANT:  NOLLING, JORK
;  APPLICANT:  ZENG, QIANDONG
;  APPLICANT:  GREENE, JONATHAN R.
;  TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
;  TITLE OF INVENTION:  AND USES THEREFOR
;  FILE REFERENCE: 2976-4031
;  CURRENT APPLICATION NUMBER: US/10/179,131
;  CURRENT FILING DATE:  2002-06-21
;  NUMBER OF SEQ ID NOS: 10194
; SEQ ID NO 7998
;   LENGTH: 1182
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-10-179-131-7998

  Query Match            57.6%;  Score 3739;  DB 25;  Length 1182;
  Best Local Similarity  62.2%;  Pred. No. 1.7e-263;
  Matches  770;  Conservative  119;  Mismatches  227;  Indels  122;  Gaps   24;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              || || || | :|:|:|| ||||||:||:||||: | |||:|||  ||||||||||||||
Db          9 MLLQFLLLSLCVSVATAKVITGVFNSFDSLTWTRAGNYAYKGPNRPTWNAVLGWSLDGTS 68

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              |||||||||||||||: | ||||||||||||||||||||||||||||:| |||:: | |
Db         69 ANPGDTFTLNMPCVFKFITDQTSVDLTADGVKYATCQFYSGEEFTTFSSLKCTVSNTLTS 128

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              |||| |||||||:||||||||  ||| |||| |||||||||||| |||||:|||:  |
Db        129 SIKALGTVTLPISFNVGGTGSLVDLESSKCFKAGTNTVTFNDGDKKISIDVDFEKTNEDA 188

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              | |  |||::||:|||:  :||||| |||||| |||:  :||  ||||:|:||||||||
Db        189 SGYFIASRLIPSINKASITYVAPQCANGYTSGAMGFTIGSGDTTIDCSNVHVGITKGLND 248

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||:||||:| || |||:| || | |:|||||||| | | | :| |:  | ||||| | |
Db        249 WNFPVSSDSLSYNKTCSSTGISITYENVPAGYRPFFDVYTSVSDQNR-QLRYTNDYACVG 307
```

```
Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
Db        308 SSLQSKPFNLRLRGYNNSEANSNGFVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 367

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
Db        368 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 427

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
Db        428 TDSIDTVVVQVPSPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYW 487

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
Db        488 SQSYATSSTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIRE 547

Qy        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
Db        548 PPNPTVTTTEYWSQSFATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYVTTSTITA 607

Qy        601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
Db        608 PPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYW 667

Qy        661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
Db        668 SQSYATTTTVTGPPGGTDTVIIREPPNQTVTTTEYWSQSYATTTTVTAPPGGTDTVIIRE 727

Qy        721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780
Db        728 PPNPTVTTTEYWSQSYATTTTVTGPPGGTDTVIIREPPNPTVTTTEYWSQSYATTTTVTA 787

Qy        781 PPGGTDTVIIYE----SMSSSKISTSSNDITSIIPS---------FSRPHYVNSTTSDLS 827
Db        788 PPGGTATVIIREPPNPTVTTTEYWSQSYATTTTVTGPPGGTDTVIIREPPNPTVTTTE-- 845

Qy        828 TFESSSMNTPTSISS--DGMLLSSTTLVTE----SETTTELICSDGKECSRLSSSSGIVT 881
Db        846 -YWSQSYATTTTVTAPPGG---TATVIIREPPNYTVTTTEY-------WSQSYATTTTVT 894

Qy        882 NPDSNESSIV-----TSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQ 936
Db        895 GPPGGTDTVIIREPPSPTVTTTEYWSQSYATTTTVTAPPGG-----------TATVIIR 942

Qy        937 TTPNPL-------SSSVTSLTQLSSIPSVSES-------ESKVTFTSNGDNQSGTHDSQS 982
Db        943 EPPNPTVTTTEYWSQSYATTTTVTGPPGGTDTVIIREPPNPTVTTTE--------YWSQS 994

Qy        983 TSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTT---SNSITEDITTSQPT 1039
Db        995 YAT--------TTTVTAPPGGTATVIIREPPN-----YTVTTTEYWSQSYATTTTVTGPP 1041

Qy       1040 GDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTST----E 1095
Db       1042 GGTDTVIIREPPSPTVTTT--------EYWSQSYATTTTVTAPPGG------TATVIIWE 1087

Qy       1096 IEATTTSPTEAPSPAVSSGTDVTTEP--TDT---REQPT-TLSTTSKTNSELVATTQATN 1149
Db       1088 PPNPTVTTTEYWSQSYATTTTVTGPPGGTDTVIIREPPSPTVTTTEYWSQSYATTTTVTA 1147

Qy       1150 ENGG-------KSPSTDLTSSLTTGTSASTSANSELVT 1180
Db       1148 PPGGTATVIIREPPNYTVT---TTEYWSQSYATTTTVT 1182
```

RESULT 13
US-10-179-131-10126
; Sequence 10126, Application US/10179131
; GENERAL INFORMATION:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHIMER JR., GEORGE H.
; APPLICANT: KESSLER, MARCO
; APPLICANT: NOLLING, JORK
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: 2976-4031
; CURRENT APPLICATION NUMBER: US/10/179,131
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 10194
; SEQ ID NO 10126
; LENGTH: 1581
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-179-131-10126

Query Match 57.5%; Score 3736.5; DB 25; Length 1581;
Best Local Similarity 58.4%; Pred. No. 4e-263;
Matches 786; Conservative 128; Mismatches 264; Indels 167; Gaps 29;

```
Qy          7 LLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDT 66
Db         13 VIHIFDSVDCKNFITGVFNSFDSLTWTRAGNYAYKGPNRPTWNAVLGWSLDGTSANPGDT 72

Qy         67 FTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFG 126
Db         73 FTLNMPCVFKFITDQTSVDLTAEGVKYATCQFYSGEEFTTFSSLKCTVSNTLTSSIKALG 132

Qy        127 TVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYA 186
Db        133 TVTLPISFNVGGTGSSVDLESSQCFKAGTNTVTFNDGDKKISIDVDFEKTNEDASGYFIA 192

Qy        187 SRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVS 246
Db        193 SRLIPSINKVSITYVAPQCANGYTSGAMGFIVLTGDTTIDCSNVHVGITKGLNDWNFPVS 252

Qy        247 SESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSK 306
Db        253 SDSLSYNKTCSSTGISITYENVPAGYRPFFDVYTSVSGQNR-QLRYTNDYACVGSSLQSK 311

Qy        307 PFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTT 366
Db        312 PFNLRLRGYNNSEANSNGFVIVATTRTVTDSTTAVTTLPFNPSIDKTKTIEILQPIPTTT 371

Qy        367 ITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 426
Db        372 ITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDT 431

Qy        427 VVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFAT 486
Db        432 VVVQVPSPNPTVSTTEYWSQSYATTTTITAPPGGTDTVIIREPPNHTVTTTEYWSQSYAT 491

Qy        487 TTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTV 546
Db        492 TTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTITAPPGGTDTVIIREPPNHTV 551

Qy        547 TTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTD 606
Db        552 TTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYTTTTTVTAPPGGTD 611

Qy        607 TVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYAT 666
Db        612 SVIIREPPNPTVTTTEYWSQSYATTTTVTATPGGTDTVIIREPPNYTVTTTEYWSQSYAT 671

Qy        667 TTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTV 726
Db        672 TTTVTAPPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNYTV 731

Qy        727 TTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTD 786
Db        732 TTTEYWSQSYATTSTITAPPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTGPPGGTD 791

Qy        787 TVIIYE--------------SMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESS 832
```

```
Db        792 TVIIREPPSPTVTTTEYWSQSYATTTTVTAPPGGTATVIIKEPPNYTVTTTE----YWSQ 847

Qy        833 SMNTPTSISS--DGMLLSSTTLVTE----SETTTELICSDGKECSRLSSSSGIVTNPDSN 886

Db        848 SYATTTTVTGPPGG---TDTVIIREPPNPTVTTTEY-------WSQSFATTTTVTAPPGG 897

Qy        887 ESSIV-----TSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNP 941

Db        898 TDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGG-----------TDSVIIREPPNP 945

Qy        942 L-------SSSVTSLTQLSSIPSVSES-------ESKVTFTSNGDNQSGTHDSQSTSTEI 987

Db        946 TVTTTEYWSQSYTTTTTVTAPPGGTDSVIIREPPNPTVTTTE--------YWSQSYATTT 997

Qy        988 EIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTT---SNSITEDITTSQPTGDNGD 1044

Db        998 TVTATPG--------GTDTVIIREPPN-----YTVTTTEYWSQSYATTTTVTAPPGGTDS 1044

Qy       1045 NTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGL--------TTSTEI 1096

Db       1045 VIIREPPNYTVTTT--------EYWSQSYATTTTVTAPPGGTDSGIIREPPNSTVTTTEY 1096

Qy       1097 ----EATTTSPT------------EAPSPAVSS----------GTDVTTEP--TDT---R 1125

Db       1097 WSQSYATTTTVTAPPGGTDSVIIREPPDPTVTSTEYWSQSFATTTTVTAPPGGTDSVIIR 1156

Qy       1126 EQPT-TLSTTSKTNSELVATTQATNENGG-------KSPSTDLTSSLTTGTSASTSANSE 1177

Db       1157 EPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVT---TTEYWSQSFATTT 1213

Qy       1178 LVTS--GSVTGGAVASASN---------DQSHSTSVT------NSNSIVSNTP--QTTLS 1218

Db       1214 TVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTT 1273

Qy       1219 QQVTSSSPSTNTFIASTYDGSGSII 1243

Db       1274 TEYWSQSYATTTTVTAPPGGTATVI 1298
```

RESULT 14
US-60-385-568-214
; Sequence 214, Application US/60385568
; GENERAL INFORMATION:
; APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Giandong Zeng
; TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and Their Use
; FILE REFERENCE: 032796-134
; CURRENT APPLICATION NUMBER: US/60/385,568
; CURRENT FILING DATE: 2002-06-05
; PRIOR FILING DATE: June 5, 2002
; NUMBER OF SEQ ID NOS: 418
; SEQ ID NO 214
; LENGTH: 1523
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-385-568-214

Query Match 57.2%; Score 3716.5; DB 27; Length 1523;
Best Local Similarity 52.0%; Pred. No. 1.1e-261;
Matches 798; Conservative 148; Mismatches 300; Indels 289; Gaps 30;

```
Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60

Db          1 MLLQFLLLSLCVSVATAKVITGIFDSFNSLTWTNAASYSYRGPANPTWTAVIGWSLDGAT 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120

Db         61 ASAGDTFTLDMPCVFKFITDQTSIDLVADGRTYATCNLNSAEEFTTFSSVSCTVTTTMTA 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180

Db        121 DTKAIGTVTLPFSFSVGGSGSDVDLANSQCFTAGINTVTFNDGDTSISATVDFEKSTVAS 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240

Db        181 SDRILLSRILPSLSQAVSLFLPQECANGYTSGTMGFSTAGTGATIDCSTVHVGISNGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300

Db        241 WNYPISSESFSYTKTCTSTSVLVTYQNVPAGYRPFVDAYVSATRVSSYAMRYTNIYACVG 300

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

Db        301 AASVDDSFTHTWSGYSNSQAGSNGITIVVTTRTVTDSTTAVTTLPFNSESDKTKTIEILQ 360

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Db        361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480

Db        421 TDSIDTVVVQVPSPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYW 480

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540

Db        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540

Qy        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600

Db        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTA 600

Qy        601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660

Db        601 PPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYW 660

Qy        661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTV-------------- 706

Db        661 SQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIRE 720

Qy        707 ----------------------TAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTA 744

Db        721 PPNPTVTTTEYWSQSYVTTSTITAPPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTA 780

Qy        745 PPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYE----SMSSSKIS 800

Db        781 PPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNYTVTTTEYW 840

Qy        801 TSSNDITSIIPS---------FSRPHYVNSTTSDLSTFESSSMNTPTSIS-----SDGML 846

Db        841 SQSYATTTTVTAPPGGTDTVIIREPPSPTVTTTE---YWSQSYATTTTVTAPPGGTDTVI 897

Qy        847 L------------------SSTTLVTESETTTELI---------CSDGKECSRLSSSSGI 879

Db        898 IREPPNPTVTTTEYWSQSYATTTTVTSPPGGTDIVIIREPPNPTVTTTEYWSQSYATTTT 957

Qy        880 VTNPDSNESSIV-----TSTVPTASTMSDSLSSTDGISA------------------TSS 916

Db        958 VTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTT 1017

Qy        917 DNVSKSGVSVTTET-----------------SVTT------------------------ 934

Db       1018 EYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVI 1077

Qy        935 IQTTPNPL-------SSSVTSLTQLSSIPSVSE------------------SES-----K 964

Db       1078 IREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSYATTLT 1137

Qy        965 VTFTSNGDN------QSGTHDSQSTST-----------EIEIVTTSSTKVLPPVVSSNT- 1006

Db       1138 ITAPPGGTNSVIIRVHSSTNDESSESTFSTLSVPSFSGSISIVSTVSR---PHYVNSTVT 1194

Qy       1007 ---DLTSEPTNTREQPTTLSTTSNSITE----------DITTSQPTGDNGDNTS----- 1047

Db       1195 HLPSSSSKPVDIPSSDVVPSTNDNSLTSLTGSENGETSVAISTTFCDDENGCQTSIPQGS 1254

Qy       1048 -STNPVPTVATST-LASASEEDNKSGSHESAST-------------SLK-------PSMG 1085
```

```
Db       1255 VVRTTA  ATPTPIIGDNNGSGKSKSGELSSTGSVTTNTATPDVPSTKVPSNPGAPGTG 1314

Qy       1086 E       PEIEATTTSPTEAPSPAVSSGTDVTTEPTDT--------------REQPTTL 1131

Db       1315         PSTETQTTNNVPGSPNIPATGT-TDIIRESTTVSHTVTGNGNTGVPMNPNPALT 1373

Qy       11           TNSELVATTQATNE------NGGKSPSTDLTS--SLTTGTSASTSANSELVTSGS 1183

Db       1            SLTG----ATNSATNPSHETGVNTGSRGSTNIVTPPSSATATVVIPGTDNGATTKGQ 1429

Qy       1184 VTGGAVAS------------------ASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSS 1224

Db       1430 DTAGGNSNGFTATTNTQGGNNEPGNQPGTNTTGEPVGTTGTQSVESTSQPTTLSQQTTSS 1489

Qy       1225 SPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLF 1259

Db       1490 LISTP--LASTFDGSGSIVQHSGWLYVLLTAISIF 1522
```

RESULT 15
US-60-385-568-323
; Sequence 323, Application US/60385568
; GENERAL INFORMATION:
; APPLICANT: Deborah Willins, Arturo Morales, Guillaume Cottarel, Giandong Zeng
; TITLE OF INVENTION: Identification of Candida Cell Surface Proteins and Their Use
; FILE REFERENCE: 032796-134
; CURRENT APPLICATION NUMBER: US/60/385,568
; CURRENT FILING DATE: 2002-06-05
; PRIOR FILING DATE: June 5, 2002
; NUMBER OF SEQ ID NOS: 418
; SEQ ID NO 323
; LENGTH: 1593
; TYPE: PRT
; ORGANISM: Candida albicans
US-60-385-568-323

Query Match 55.7%; Score 3620; DB 27; Length 1593;
Best Local Similarity 53.9%; Pred. No. 1.3e-254;
Matches 763; Conservative 158; Mismatches 308; Indels 186; Gaps 25;

```
Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60

Db          1 MLLQFLLLSLCVSVATAKVITGIFDSFNSLTWTNAASYSYRGPANPTWTAVIGWSLDGAT 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120

Db         61 ASAGDTFTLDMPCVFKFITDQTSIDLVADGRTYATCNLNSAEEFTTFSSVSCTVTTTMTA 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180

Db        121 DTKAIGTVTLPFSFSVGGSGSDVDLANSQCFTAGINTVTFNDGDTSISTTVDFEKSTVAS 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240

Db        181 SDRILLSRILPSLSQAVSLFLPQECANGYTSGTMGFSTAGTGATIDCSTVHVGISNGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300

Db        241 WNYPISSESFSYTKTCTSTSVLVTYQNVPAGYRPFVDAYVSATRVSSYAMRYTNIYACVG 300

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

Db        301 AASVDDSFTHTWSGYSNSQAGSNGITIVVTTRTVTDSTTAVTTLPFNSESDKTKTIEILQ 360

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Db        361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480

Db        421 TDSIDTVVVQVPSPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYW 480

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540

Db        481 SQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIRE 540

Qy        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600

Db        541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTA 600

Qy        601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660

Db        601 PPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYW 660

Qy        661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720

Db        661 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIRE 720

Qy        721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780

Db        721 PPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTA 780

Qy        781 PPGGTDTVIIYE--------------SMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDL 826

Db        781 PPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNYTVTTTE-- 838

Qy        827 STFESSSMNTPTSIS-----SDGMLL------------------SSTTLVTESETTTELI 863

Db        839 --YWSQSYATTTTVTGPPGGTDTVIIREPPNPTVTTTEYWSQSYATTTTVTSPPGGTDIV 896

Qy        864 ---------CSDGKECSRLSSSSGIVTNPDSNESSIV-----TSTVPTASTMSDSLSSTD 909

Db        897 IIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTT 956

Qy        910 GISA------------------TSSDNVSKSGVSVTTETS------VTTIQTTPNPL--- 942

Db        957 TVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTT 1016

Qy        943 ----SSSVTSLTQLSSIPSVSE------------------SESKVTFTSNGDNQSGT--- 977

Db       1017 TEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDTV 1076

Qy        978 --------------HDSQSTSTEIEIV-----------------TTSSTKVLPPVVSSNT 1006

Db       1077 IIREPPNYTVTTTEYWSQSYATTTTVTGPPGGTDTVIIREPPNPTVTTTEYWSQSYATTT 1136

Qy       1007 DLTSEPTNT-----REQPT-TLSTT---SNSITEDITTSQPTGDNGDNTSSTNPVPTVAT 1057

Db       1137 TVTSPPGGTDIVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTT 1196

Qy       1058 STLASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDV 1117

Db       1197 T--------EYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTT--TEYWSQSYATTTTV 1246

Qy       1118 TTEP--TDT---REQPT-TLSTTSKTNSELVATTQATNENGG-------KSPSTDLT--- 1161

Db       1247 TAPPGGTDTVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTE 1306

Qy       1162 ----SSLTTGTSASTSANSELVTSGSVTGGAVASASN-DQSHSTSVT------NSNSIVS 1210

Db       1307 YWSQSYATTTTVTAPPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTAPPGGTNTVII 1366

Qy       1211 NTP--QTTLSQQVTSSSPSTNTFIASTYDGSGSII 1243

Db       1367 REPPSPTVTTTEYWSQSYATTTTVTAPPGGTATVI 1401
```

Search completed: June 11, 2003, 17:20:09
Job time : 309 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 17:13:36 ; Search time 58 Seconds
(without alignments)
2242.806 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQQFTLLFLYLSIASAKTI.....SIIQHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	939	14.5	1537	10	US-09-801-368-104	Sequence 104, App
2	932.5	14.4	1322	10	US-09-801-368-114	Sequence 114, App
3	721.5	11.1	5179	9	US-10-025-380-1068	Sequence 1068, Ap
4	721.5	11.1	5179	10	US-09-922-217-1068	Sequence 1068, Ap
5	721.5	11.1	5179	10	US-09-833-263-1068	Sequence 1068, Ap
6	711	10.9	1075	10	US-09-801-368-110	Sequence 110, App
7	675.5	10.4	1367	10	US-09-801-368-108	Sequence 108, App
8	640.5	9.9	2344	10	US-09-815-242-12713	Sequence 12713, A
9	545.5	8.4	957	9	US-10-025-380-1065	Sequence 1065, Ap
10	545.5	8.4	957	10	US-09-922-217-1065	Sequence 1065, Ap
11	545.5	8.4	957	10	US-09-833-263-1065	Sequence 1065, Ap
12	522.5	8.0	5877	9	US-10-142-515-11	Sequence 11, Appl
13	501	7.7	1169	10	US-09-801-368-106	Sequence 106, App
14	463	7.1	688	10	US-09-864-761-36047	Sequence 36047, A
15	428	6.6	1236	9	US-09-769-787-109	Sequence 109, App
16	417	6.4	560	10	US-09-815-242-13057	Sequence 13057, A
17	392	6.0	502	10	US-09-815-242-5904	Sequence 5904, Ap
18	390	6.0	596	9	US-10-063-547-100	Sequence 100, App
19	390	6.0	596	9	US-10-174-590-310	Sequence 310, App
20	390	6.0	596	9	US-10-176-758-310	Sequence 310, App
21	390	6.0	596	9	US-10-063-616-100	Sequence 100, App
22	390	6.0	596	9	US-10-175-737-310	Sequence 310, App
23	390	6.0	596	9	US-10-063-502-100	Sequence 100, App
24	390	6.0	596	9	US-10-173-706-310	Sequence 310, App
25	390	6.0	596	9	US-10-175-738-310	Sequence 310, App
26	390	6.0	596	9	US-10-175-752-310	Sequence 310, App
27	390	6.0	596	9	US-10-176-482-310	Sequence 310, App
28	390	6.0	596	9	US-10-176-757-310	Sequence 310, App
29	390	6.0	596	9	US-10-176-913-310	Sequence 310, App
30	390	6.0	596	9	US-10-180-552-310	Sequence 310, App
31	390	6.0	596	9	US-10-180-557-310	Sequence 310, App
32	390	6.0	596	9	US-10-173-700-310	Sequence 310, App
33	390	6.0	596	9	US-10-174-572-310	Sequence 310, App
34	390	6.0	596	9	US-10-174-579-310	Sequence 310, App
35	390	6.0	596	9	US-10-174-582-310	Sequence 310, App
36	390	6.0	596	9	US-10-174-588-310	Sequence 310, App
37	390	6.0	596	9	US-10-175-739-310	Sequence 310, App
38	390	6.0	596	9	US-10-175-740-310	Sequence 310, App
39	390	6.0	596	9	US-10-175-743-310	Sequence 310, App
40	390	6.0	596	9	US-10-176-488-310	Sequence 310, App
41	390	6.0	596	9	US-10-176-492-310	Sequence 310, App
42	390	6.0	596	9	US-10-176-747-310	Sequence 310, App
43	390	6.0	596	9	US-10-176-750-310	Sequence 310, App
44	390	6.0	596	9	US-10-176-985-310	Sequence 310, App
45	390	6.0	596	9	US-10-176-987-310	Sequence 310, App

ALIGNMENTS

RESULT 1
US-09-801-368-104
; Sequence 104, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 104
; LENGTH: 1537
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-104

Query Match 14.5%; Score 939; DB 10; Length 1537;
Best Local Similarity 25.8%; Pred. No. 1.1e-37;
Matches 414; Conservative 246; Mismatches 520; Indels 426; Gaps 84;

```
Qy          2 LQQFTLLFLYLSIASAKT---------ITGVFDSF------NSLTWSNAANYAF------ 40
              |  |||| |  |:||  |          :|:  :|      :| |:|||| |:
Db         10 LAVFTLLAL-TSVASGATEACLPAGQRKSGMNINFYQYSLKDSSTYSNAAYMAYGYASKT 68
```

```
Qy          41 ----------------------------KGPGYPTW-----------NAVLGWSLD--G 58
Db          69 KLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGACSNSQGIAYWSTDLFG 128

Qy          59 TSANPGDTFTLNM----------PCVFKYTTSQ----------------------TSVD 85
Db         129 FYTTPTNV-TLEMTGYFLPPQTGSYTFKFATVDDSAILSVGGATAFNCCAQQQPPITSTN 187

Qy          86 LTADGVK----------YATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGT----VTLP 131
Db         188 FTIDGIKPWGGGSLPPNIEGTVYMYAGYYYPM---------KVVYSNAVSWGTLPISVTLP 238

Qy         132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV-DPSAYLYASRVM 190
Db         239 -----DGTTVSDDFEG--------YVYSFDD-------DLSQSNCTVPDPSNYA-VSTTT 277

Qy         191 PSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESF 250
Db         278 TTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTTASTIITTTEPWNSTFTSTST 337

Qy         251 SYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTL 310
Db         338 ELTTVTGTNGVRTDETII------VIRTPTTATTAITTTEPWNSTFTSTSTEL------- 384

Qy         311 RWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVTTL-PFNPSVDKTKT---------- 355
Db         385 ------TTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTVTGTNGL 438

Qy         356 ------IEILQPIPTTTITT-------SYVGVTTSYLTKTA----PIGETATVIVDVPYH 398
Db         439 PTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTVTGTNGLPTDET-IIVIRTPTT 497

Qy         399 TTT--TVTSEWTGTIT------TTTTRTN--PTDSIDTVVVQVPLPNPTVSTT-----EY 443
Db         498 ATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTD--ETIIV---IRTPTTATTAMTTPQP 552

Qy         444 WSQSFATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTV 490
Db         553 WNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTEPWNSTFTSTSTEMTTV 612

Qy         491 TAPPG-GTDS--VIIREPPNPT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VII 538
Db         613 TGTNGLPTDETIIVIRTPTTATTAITTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVI 672

Qy         539 REPPNPT--VTTTEYWSQSYATT----TTVTAPPG-GTDS--VIIREPPNHT--VTTTEY 587
Db         673 RTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTTGLPTDETIIVIRTPTTATTAMTTTQP 732

Qy         588 WSQSYATT----TTVTAPPG-GTD--TVIIREPPNHTV--TTTEYWSQSFATT----TTV 634
Db         733 WNDTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTV 792

Qy         635 T---GPPSGTDTVIIREPPNP--TVTTTEYWSQSYATT----TTITAP---PGETDTVLI 682
Db         793 TGTNGQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMTTITGTNGVPTDETVIVI 852

Qy         683 REPPNHTV--TTTEYWSQSYATT----TTVTAPPGE-TD--TVLIREPPNHTV--TTTEY 731
Db         853 RTPTSEGLISTTTEPWTGTFTSTSTEMTTITGTNGQPTDETVIVIRTPTSEGLISTTTEP 912

Qy         732 WSQSYATTTTVTAPPGGTD-------TVIIREPPNPTV--TTTEYWSQSFATTTTVTAPP 782
Db         913 WTGTFTSTSTEMTHVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEVTTI 972

Qy         783 GGT------DTVIIYESMSSSKI------------STSSNDITSIIPSFSRP-------- 816
Db         973 TGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVI 1032

Qy         817 ------HYVNSTTSD-LSTFESSSMNTPTSISSDGMLLSSTTLVTESETT----TELICS 865
Db        1033 RTPTSEGLVTTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSS 1092

Qy         866 DGKECSRLSSSSGIVTNP--DSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSG 923
```

```
Db        1093 SGQITSSITSSRPIIT-PFYPSNGTSVISSSVISSSVTSSLFTSSPVISSSVISSSTTTS 1151

Qy         924 VSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQST 983
Db        1152 TSIFSESSKSSV--IPTSSSTSGSSESETSSAGSVSSS----SFISSESSKSPTYSSSS- 1204

Qy         984 STEIEIVTTSST-----KVLPPVVSSNTDLTSEPTNTREQPTTLSTTS-------NSITE 1031
Db        1205 ---LPLVTSATTSQETASSLPPATT---------TKTSEQTTLVTVTSCESHVCTESISP 1252

Qy        1032 DITTSQPTGDNGDNTSSTN--PVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSG 1089
Db        1253 AIVSTATVTVSGVTTEYTTWCPISTTETTKQTKGTTEQTTETTKQTTVVTISSCESDVCS 1312

Qy        1090 LTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTT-------SKTNSELV 1142
Db        1313 KTASPAIVSTSTATINGVTTEYTTWCPIST--TESRQQTTLVTVTSCESGVCSETASPAI 1370

Qy        1143 ATT-----------------QATNENGGKSPSTDLTSSLTTGT-SASTSANSELVTSGSV 1184
Db        1371 VSTATATVNDVVTVYPTWRPQTANEESVSSKMNSATGETTTNTLAAETTTNTVAAETITN 1430

Qy        1185 TGGA-------------------VASASNDQSHSTSV-----TNSNSIVSNTPQTTLSQQ 1220
Db        1431 TGAAETKTVVTSSLSRSNHAETQTASATDVIGHSSSVVSVSETGNTKSLTSSGLSTMSQQ 1490

Qy        1221 VTSSSP------STNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db        1491 PRSTPASSMVGYSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1536
```

RESULT 2
US-09-801-368-114
; Sequence 114, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 114
; LENGTH: 1322
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-114

Query Match 14.4%; Score 932.5; DB 10; Length 1322;
Best Local Similarity 28.1%; Pred. No. 1.9e-37;
Matches 400; Conservative 215; Mismatches 468; Indels 339; Gaps 79;

```
Qy          28 NSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKYTTSQ----S 83
Db          50 DSSTYSNAAYMAY---GYASKT-----KLGSVGGQTDISIDYNIPCV---SSSGTFPCPQ 98
```

```
Qy          84 VDLTAD-GVK-YATCQ------FYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIA-- 133
Db          99 EDLYGNWGCKGIGACSNNPIIAYWSTDLFGFYTTPT--------------NVTLEMTGY 143

Qy         134 FNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSL 193
Db         144 FLPPQTGSYT----------------------------FKFATVDDSAIL--------- 165

Qy         194 NKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDC----------SNIHIGITKGLNDWN- 242
Db         166 ------------------------SVGGSIAFECCAQEQPPITSTNFTI---NGIKPWNG 198

Qy         243 -YPVSSESFSYTKTCTSNGIQIKYQNVPA-GYRPF-------------IDAYISATDVN- 286
Db         199 SPPDNITGTVYMYAGFYYPMKIVYSNAVAWGTLPISVTLPDGTTVSDDFEGYVYTFDNNL 258

Qy         287 ---QYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDA-------GSNGI-----VIVATT 331
Db         259 SQPNCTIPDPSNYTVSTTITTTEP----WTGTFTSTSTEMTTVTGTNGVPTDETVIVIRT 314

Qy         332 RTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTIT-TSYVGVTTSYLTKTAPIGETAT 390
Db         315 PT-TASTIITTTEPWNSTFTSTST-------ELTTVTGTNGVRTDETIIVIRTPTTATTA 366

Qy         391 VIVDVPYHTTTTVTSEWTGTITTTTTRTN--PTDSIDTVVVQVPLPNPT-VSTTEYWSQS 447
Db         367 ITTTEPWNSTFTSTS----TELTTVTGTNGLPTDE-TIIVIRTPTTATTAMTTTQPWNDT 421

Qy         448 FATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTVTAPP 494
Db         422 FTSTSTELTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTVTGTN 481

Qy         495 G-GTDS--VIIREPPNPT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPP 542
Db         482 GLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTDETIIVIRTPT 541

Qy         543 NPT--VTTTEYWSQSYATT----TTVTAPPG-GTDS--VIIREPPNHT--VTTTEYWSQS 591
Db         542 TATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTEPWNST 601

Qy         592 YATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTVTGP- 637
Db         602 FTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTQPWNDTFTSTSTEMTTVTGTN 661

Qy         638 --PSGTDTVIIREPPNPT--VTTTEYWSQSYATT----TTITAP---PGETDTVLIREPP 686
Db         662 GLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTDETIIVIRTPT 721

Qy         687 NHT--VTTTEYWSQSYATT----TTVTAP---PGETDTVLIREPPNHTV--TTTEYWSQS 735
Db         722 TATTAMTTTQPWNDTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGT 781

Qy         736 YATT----TTVTAPPG-GTD--TVIIREPPNP--TVTTTEYWSQSFATT----TTVTAPP 782
Db         782 FTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMTTITGTN 841

Qy         783 G--GTDTVIIYESMS---SSKISTSSNDITSIIPS---FSRPHYVNSTTSDLSTFESSSM 834
Db         842 GQPTDETVIIVKTPTTAISSSLSSSSGQITSFITSARPIITPFYPSNGTSVISSSVISSS 901

Qy         835 NTPTSISSDGM---LLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIV 891
Db         902 DTSSLVISSSVTSSLVTSSPVISSSFISSPVISSTTTSASILSESS------------K 948

Qy         892 TSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQ 951
Db         949 SSVIPTSSSTSGSSESETGSASSAS---SSSSISSESPKSTYSSSSLP-PVTSATTSQEI 1004

Qy         952 LSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSE 1011
Db        1005 TSSLPPVTTTKT-----------------SEQTTLVTVTSCESHVCTESISSAIVSTAT 1046

Qy        1012 PT---NTREQPTTLSTTSNSITEDIT-TSQPTGDNGDNTSSTNPVPTVATSTLASASEED 1067
Db        1047 VTVSGATTEYTTWCPISTTEITKQTTETTKQTKGTTEQTTETTKQTTVVT---ISSCESD 1103

Qy        1068 --NKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTR 1125
Db        1104 VCSKTASPAIVSTSTATINGVTTEYTTWCPISTTESKQQTTLVTVTSCGSGVCSETT--- 1160

Qy        1126 EQPTTLSTTSKTNSELVAT-----TQATNENGGKSPSTDLTSSLTTGTSAST------SA 1174
Db        1161 -SPAIVSTATATVNDVVTVYSTWRPQTTNEQSVSSKMNSATSETTTNTGAAETTTSTGAA 1219

Qy        1175 NSELVTSGSVT-----GGAVASASNDQSHSTSV-----TNSNSIVSNTPQTTLSQQVTSS 1224
Db        1220 ETKTVVTSSISRFNHAETQTASATDVIGHSSSVVSVSETGNTKSLTSSGLSTMSQQPRST 1279

Qy        1225 ------SPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db        1280 PASSMVGSSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1321
```

RESULT 3
US-10-025-380-1068
; Sequence 1068, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1068

Query Match 11.1%; Score 721.5; DB 9; Length 5179;
Best Local Similarity 25.3%; Pred. No. 1.2e-26;
Matches 335; Conservative 158; Mismatches 554; Indels 279; Gaps 54;

```
Qy          59 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 118
Db        1647 TTTTPPPTTTPSPP-----TTTPSPPITTTTTPPPTT--------TPSSPITTTPSPPT 1693

Qy         119 KSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV 178
Db        1694 TTMTTPSPTTTPSSPITTTTTPSSTTTPSPPPTTMTTPSPTTTPSPPTTTM------TTL 1747

Qy         179 DP---SAYLYASRVMPSLNKVT---------TLFVAPQCE-NGY-TSGTMGFSSSNGDVA 224
Db        1748 PPTTTSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVPLCNWTGWLDSGKPNFHKPGGDTE 1807

Qy         225 IDCSNIHIGITKGLNDWNYPVSSESFSY----------TKTC-TSNGIQIKYQN-VPAGY 272
```

```
Db       1808 L------IGDVCG-PGWAANISCRATMYPDVPIGQLGQTVVCDVSVGLICKNEDQKPGGV 1860

Qy        273 RPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTR 332

Db       1861 IPM----------AFCLNYEINVQCC--ECVTQPTTMTTTTTENPTPPTTTPITTTTTV 1907

Qy        333 TVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIG-ETATV 391

Db       1908 TPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTT 1967

Qy        392 IVDVPYHTTTTVTSEWTGT---------ITTTT----------TRTNPTDSIDTVVVQVP 432

Db       1968 ---TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTP 2024

Qy        433 LPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTA 492

Db       2025 TPTPTGTQTP-TTTPITTTTTVTPTPTPTGT---QTPTTTPITTT--------TTVTPTP 2072

Qy        493 PPGGTD-------SVIIREPPNPTVTTTEYWSQSFATTTTVTAP---PGGTD-------S 535

Db       2073 TPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPIT 2132

Qy        536 VIIREPPNPTVTTTEYWSQSYATTTTVTAP---PGGTDSVIIREPPNHTVTTTEYWSQSY 592

Db       2133 TTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-----PTTTPITTT------- 2180

Qy        593 ATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTD---------T 643

Db       2181 -TTVTPTPTPTGTQT-----PTTTPITTT--------TTVTPTPTPTGTQTPTTTPITTT 2226

Qy        644 VIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATT 703

Db       2227 TTVTPTPTPTGTQTP-TTTPITTTTTVTPTPTPTGT--------QTPTTT-----PITTT 2272

Qy        704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTD---------TVII 754

Db       2273 TTVTPTPTPTGT---QTPTTTPITTT--------TTVTPTPTPTGTQTPTTTPITTTTTV 2321

Qy        755 REPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFS 814

Db       2322 TPTPTPTGTQTP-TTTPITTTTTTVTPTPTPTGT----QTPTTTPITTT----TTVTPTPT 2372

Qy        815 RPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLS 874

Db       2373 PTGTQTPTTTPITT---TTTVTPTP-TPTGTQTPTTTPITTTTTVTPTPTPTGTQ----- 2423

Qy        875 SSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATS--SDNVSKSGVSVTTETSV 932

Db       2424 -------TPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 2476

Qy        933 TTIQT---------TPNPLSSSVTSLTQLSSIPSVSESESKVT---FTSNGDNQSGTHDS 980

Db       2477 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTG 2536

Qy        981 QSTSTEIEIVTTSSTKVLP-----------PVVSSNT-DLTSEPTNTREQPTTLSTTSNS 1028

Db       2537 TQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTT 2596

Qy       1029 ITEDITTSQPTGDNGDNTSSTNPVPTVATST-LASASEEDNKSGSHESASTSLKPSMGEN 1087

Db       2597 VT---PTPTPT---GTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPT 2650

Qy       1088 SGLTTSTEIEATTTSPTEAPSPA---------VSSGTDVTTEPTDTREQ----------- 1127

Db       2651 GTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTT 2710

Qy       1128 -----PTTLSTTSKTNSELVATTQAT---NENGGKSPSTDLTSSLTTGTSASTSANSELV 1179

Db       2711 TVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTP 2770

Qy       1180 TSGSVTGGAVA----SASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIAST 1235

Db       2771 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPT 2830

Qy       1236 YDGSGS 1241

Db       2831 PTPTGT 2836


RESULT 4
US-09-922-217-1068
; Sequence 1068, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole Lynn
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C13
;  CURRENT APPLICATION NUMBER: US/09/922,217
;  CURRENT FILING DATE:  2001-08-03
;  NUMBER OF SEQ ID NOS: 1124
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1068
;   LENGTH: 5179
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-922-217-1068

  Query Match             11.1%;  Score 721.5;  DB 10;  Length 5179;
  Best Local Similarity   25.3%;  Pred. No. 1.2e-26;
  Matches  335;  Conservative  158;  Mismatches  554;  Indels  279;  Gaps   54;

Qy          59 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 118

Db        1647 TTTTPPPTTTPSPP-----TTTTPSPPITTTTTPPPTT--------TPSSPITTTPSPPT 1693

Qy         119 KSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV 178

Db        1694 TTMTTPSPTTTPSSPITTTTTPSSTTTPSPPPTTMTTPSPTTTPSPPTTTM------TTL 1747

Qy         179 DP---SAYLYASRVMPSLNKVT---------TLFVAPQCE-NGY-TSGTMGFSSSNGDVA 224

Db        1748 PPTTTSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVPLCNWTGWLDSGKPNFHKPGGDTE 1807

Qy         225 IDCSNIHIGITKGLNDWNYPVSSESFSY----------TKTC-TSNGIQIKYQN-VPAGY 272

Db        1808 L------IGDVCG-PGWAANISCRATMYPDVPIGQLGQTVVCDVSVGLICKNEDQKPGGV 1860

Qy         273 RPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTR 332

Db        1861 IPM----------AFCLNYEINVQCC--ECVTQPTTMTTTTTENPTPPTTTPITTTTTV 1907

Qy         333 TVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIG-ETATV 391

Db        1908 TPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTT 1967

Qy         392 IVDVPYHTTTTVTSEWTGT---------ITTTT----------TRTNPTDSIDTVVVQVP 432

Db        1968 ---TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTP 2024

Qy         433 LPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTA 492

Db        2025 TPTPTGTQTP-TTTPITTTTTVTPTPTPTGT---QTPTTTPITTT--------TTVTPTP 2072

Qy         493 PPGGTD-------SVIIREPPNPTVTTTEYWSQSFATTTTVTAP---PGGTD-------S 535
```

```
Db       2073 TPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTTVTPTPTPTGTQTPTTTPIT 2132

Qy        536 VIIREPPNPTVTTTEYWSQSYATTTTVTAP---PGGTDSVIIREPPNHTVTTTEYWSQSY 592

Db       2133 TTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-----PTTTPITTT------- 2180

Qy        593 ATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTD---------T 643

Db       2181 -TTVTPTPTPTGTQT-----PTTTPITTT--------TTVTPTPTPTGTQTPTTTPITTT 2226

Qy        644 VIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATT 703

Db       2227 TTVTPTPTPTGTQTP-TTTPITTTTTVTPTPTPTGT--------QTPTTT-----PITTT 2272

Qy        704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTD---------TVII 754

Db       2273 TTVTPTPTPTGT---QTPTTTPITTT--------TTVTPTPTPTGTQTPTTTPITTTTTV 2321

Qy        755 REPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFS 814

Db       2322 TPTPTPTGTQTP-TTTPITTTTTVTPTPTPTGT----QTPTTTPITTT----TTVTPTPT 2372

Qy        815 RPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLS 874

Db       2373 PTGTQTPTTTPITT---TTTVTPTP-TPTGTQTPTTTPITTTTTVTPTPTPTGTQ----- 2423

Qy        875 SSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATS--SDNVSKSGVSVTTETSV 932

Db       2424 -------TPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 2476

Qy        933 TTIQT---------TPNPLSSSVTSLTQLSSIPSVSESESKVT---FTSNGDNQSGTHDS 980

Db       2477 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTG 2536

Qy        981 QSTSTEIEIVTTSSTKVLP-----------PVVSSNT-DLTSEPTNTREQPTTLSTTSNS 1028

Db       2537 TQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTT 2596

Qy       1029 ITEDITTSQPTGDNGDNTSSTNPVPTVATST-LASASEEDNKSGSHESASTSLKPSMGEN 1087

Db       2597 VT---PTPTPT---GTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPT 2650

Qy       1088 SGLTTSTEIEATTTSPTEAPSPA---------VSSGTDVTTEPTDTREQ----------- 1127

Db       2651 GTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTT 2710

Qy       1128 -----PTTLSTTSKTNSELVATTQAT---NENGGKSPSTDLTSSLTTGTSASTSANSELV 1179

Db       2711 TVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTP 2770

Qy       1180 TSGSVTGGAVA----SASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIAST 1235

Db       2771 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPT 2830

Qy       1236 YDGSGS 1241

Db       2831 PTPTGT 2836


RESULT 5
US-09-833-263-1068
; Sequence 1068, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Meagher, Madeleine J.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C12
;  CURRENT APPLICATION NUMBER: US/09/833,263
;  CURRENT FILING DATE:  2001-04-10
;  NUMBER OF SEQ ID NOS: 1093
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1068
;   LENGTH: 5179
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-833-263-1068

  Query Match             11.1%;  Score 721.5;  DB 10;  Length 5179;
  Best Local Similarity   25.3%;  Pred. No. 1.2e-26;
  Matches  335;  Conservative  158;  Mismatches  554;  Indels  279;  Gaps   54;

Qy          59 TSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDAL 118

Db        1647 TTTTPPPTTTPSPP-----TTTTPSPPITTTTTPPPTT--------TPSSPITTTPSPPT 1693

Qy         119 KSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV 178

Db        1694 TTMTTPSPTTTPSSPITTTTTPSSTTPSPPPTTMTTPSPTTTPSPPTTTM------TTL 1747

Qy         179 DP---SAYLYASRVMPSLNKVT---------TLFVAPQCE-NGY-TSGTMGFSSSNGDVA 224

Db        1748 PPTTTSSPLTTTPLPPSITPPTFSPFSTTTPTTPCVPLCNWTGWLDSGKPNFHKPGGDTE 1807

Qy         225 IDCSNIHIGITKGLNDWNYPVSSESFSY----------TKTC-TSNGIQIKYQN-VPAGY 272

Db        1808 L------IGDVCG-PGWAANISCRATMYPDVPIGQLGQTVVCDVSVGLICKNEDQKPGGV 1860

Qy         273 RPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTR 332

Db        1861 IPM----------AFCLNYEINVQCC--ECVTQPTTMTTTTTENPTPPTTTPITTTTTV 1907

Qy         333 TVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTTITTSYVGVTTSYLTKTAPIG-ETATV 391

Db        1908 TPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTT 1967

Qy         392 IVDVPYHTTTTVTSEWTGT---------ITTTT----------TRTNPTDSIDTVVVQVP 432

Db        1968 ---TPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTP 2024

Qy         433 LPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTA 492

Db        2025 TPTPTGTQTP-TTTPITTTTTVTPTPTPTGT---QTPTTTPITTT--------TTVTPTP 2072

Qy         493 PPGGTD-------SVIIREPPNPTVTTTEYWSQSFATTTTVTAP---PGGTD-------S 535

Db        2073 TPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPIT 2132

Qy         536 VIIREPPNPTVTTTEYWSQSYATTTTVTAP---PGGTDSVIIREPPNHTVTTTEYWSQSY 592

Db        2133 TTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQT-----PTTTPITTT------- 2180

Qy         593 ATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTD---------T 643

Db        2181 -TTVTPTPTPTGTQT-----PTTTPITTT--------TTVTPTPTPTGTQTPTTTPITTT 2226

Qy         644 VIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATT 703

Db        2227 TTVTPTPTPTGTQTP-TTTPITTTTTVTPTPTPTGT--------QTPTTT-----PITTT 2272

Qy         704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTD---------TVII 754

Db        2273 TTVTPTPTPTGT---QTPTTTPITTT--------TTVTPTPTPTGTQTPTTTPITTTTTV 2321

Qy         755 REPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFS 814

Db        2322 TPTPTPTGTQTP-TTTPITTTTTVTPTPTPTGT----QTPTTTPITTT----TTVTPTPT 2372

Qy         815 RPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLS 874

Db        2373 PTGTQTPTTTPITT---TTTVTPTP-TPTGTQTPTTTPITTTTTVTPTPTPTGTQ----- 2423
```

```
Qy        875 SSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATS--SDNVSKSGVSVTTETSV 932
Db       2424 -------TPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPI 2476

Qy        933 TTIQT---------TPNPLSSSVTSLTQLSSIPSVSESESKVT---FTSNGDNQSGTHDS 980
Db       2477 TTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTG 2536

Qy        981 QSTSTEIEIVTTSSTKVLP----------PVVSSNT-DLTSEPTNTREQPTTLSTTSNS 1028
Db       2537 TQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTT 2596

Qy       1029 ITEDITTSQPTGDNGDNTSSTNPVPTVATST-LASASEEDNKSGSHESASTSLKPSMGEN 1087
Db       2597 VT---PTPTPT---GTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPT 2650

Qy       1088 SGLTTSTEIEATTTSPTEAPSPA---------VSSGTDVTTEPTDTREQ----------- 1127
Db       2651 GTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTT 2710

Qy       1128 -----PTTLSTTSKTNSELVATTQAT---NENGGKSPSTDLTSSLTTGTSASTSANSELV 1179
Db       2711 TVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTP 2770

Qy       1180 TSGSVTGGAVA----SASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIAST 1235
Db       2771 TTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPTPTPTGTQTPTTTPITTTTTVTPT 2830

Qy       1236 YDGSGS 1241
Db       2831 PTPTGT 2836


RESULT 6
US-09-801-368-110
; Sequence 110, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
;  APPLICANT: Busby, Robert
;  APPLICANT:  Cali, Brian
;  APPLICANT:  Hecht, Peter
;  APPLICANT:  Holtzman, Doug
;  APPLICANT:  Madden, Kevin
;  APPLICANT:  Maxon, Mary
;  APPLICANT:  Milne, Todd
;  APPLICANT:  No. US20020128250A1man, Thea
;  APPLICANT:  Royer, John
;  APPLICANT:  Salama, Sofie
;  APPLICANT:  Sherman, Amir
;  APPLICANT:  Silva, Jeff
;  APPLICANT:  Summers, Eric
;  TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
;  FILE REFERENCE: 109272.147
;  CURRENT APPLICATION NUMBER: US/09/801,368
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: US 09/487,558
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: US 60/160,587
;  PRIOR FILING DATE: 1999-10-20
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 110
;   LENGTH: 1075
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-110

  Query Match            10.9%;  Score 711;  DB 10;  Length 1075;
  Best Local Similarity   25.9%;  Pred. No. 7.2e-27;
  Matches  324;  Conservative  192;  Mismatches  393;  Indels  342;  Gaps   66;

Qy        153 AGTNTVTFNDGDKDISIDVEF------EKSTVDPSAYL---YASRV-------------- 189
Db         23 SGATEACLPAGQRKSGMNINFYQYSLKDSSTYSNAAYMAYGYASKTKLGSVGGQTDISID 82

Qy        190 --MPSLNKVTTLFVAPQCENGYTS-GTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVS 246
Db         83 YNIPCVSSSGT-FPCPQ-EDSYGNWGCKGMGA--------CSN-----SQGIAYW----S 123

Qy        247 SESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSR---L 303
Db        124 TDLFGFYTTPTNVTLEMTGYFLPPQTGSYTFSFATVDDSAILSVGGSIAFECCAQEQPPI 183

Qy        304 QSKPFTLRWTGYKNSDAGSNGIVIVATT----------RTVTDSTTAVTTLPFNPSVDKT 353
Db        184 TSTNFTI--NGIKPWD-GSLPDNITGTVYMYAGYYYPLKVVYSNAVSWGTLPI--SVE-- 236

Qy        354 KTIEILQPIPT-TTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTT---TTVTSEWTG 409
Db        237 --------LPDGTTVSDNFEGYVYSF---DDDLSQSNCTIPDPSIHTTSTITTTTEPWTG 285

Qy        410 TIT------TTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDT 463
Db        286 TFTSTSTEMTTITDTNGQLTDETVIV---IRTPT-----------TASTIT-------- 322

Qy        464 VIIREPPNHTVTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPPNP--TVTTTE 514
Db        323 -----------TTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLITTTTE 371

Qy        515 YWSQSFATT----TTVTAPPG-GTDS--VIIREPPNP--TVTTTEYWSQSYATTTTVTAP 565
Db        372 PWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEVTT 431

Qy        566 PGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNH--TVTTTEY 623
Db        432 ITGTNG----QPTDETV------------------------IVIRTPTSEGLITTTTEP 462

Qy        624 WSQSFATT----TTVT---GPPSGTDTVIIREPPNPTV--TTTEYWSQSYATT----TTI 670
Db        463 WTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEVTTI 522

Qy        671 TAPPGE-TD--TVLIREPPNH--TVTTTEYWSQSYATT----TTVTAPPGE-TD--TVLI 718
Db        523 TGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVI 582

Qy        719 REPPNH--TVTTTEYWSQSYATTTTVTAPPGGT------DTVIIREPPNPTVTTTEYWSQ 770
Db        583 RTPTSEGLITRTTEPWTGTFTSTSTEVTTITGTNGQPTDETVIVIRTPT-TAISSSLSSS 641

Qy        771 SFATTTTVTA---------PPGGTDTV---IIYESMSSSKISTSSNDITSIIPSFSRPHY 818
Db        642 SGQITSSITSSRPIITPFYPSNGTSVISSSVISSSVTSSLVTSSSFISSSVISS------ 695

Qy        819 VNSTTSDLSTFESSSMNT--PTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSS 876
Db        696 --STTTSTSIFSESSTSSVIPTSSSTSGSSESKTSSASSSSSSSSSISSESPKSPTNSSSS 753

Qy        877 SGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATS------SDNVSKSGVSVTTET 930
Db        754 LPPVTSATTGQET--ASSLPPATTTKTS-EQTTLVTVTSCESHVCTESISSAIVSTATVT 810

Qy        931 --SVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIE 988
Db        811 VSGVTTEYTTWCPISTTETT----------------------KQTKGTTEQTKGTTEQT 847

Qy        989 IVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSS 1048
Db        848 TETTKQTTVV-TISSCESDICSKTAS----PAIVSTSTATI-----------NGVTTEY 890

Qy       1049 TNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAP- 1107
Db        891 TTWCPISTTES----------------------KQQTTLVTVTSCESGVCSETTSPA 925

Qy       1108 --SPAVSSGTDVTT-----EPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDL 1160
Db        926 IVSTATATVNDVVTVYPTWRPQTTNEQ----SVSSKMNS---ATSETTTNTGAAETKTAV 978
```

```
Qy       1161 TSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSV-----TNSNSIVSNTPQT 1215
             ||||:    | |                  |||::   ||:||     | :   ::::  :
Db        979 TSSLSRFNHAETQ--------------TASATDVIGHSSSVVSVSETGNTMSLTSSGLS 1023

Qy       1216 TLSQQVTSS------SPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
             |:|||  |:       || :   ||| || : :   : |   |  | | |
Db       1024 TMSQQPRSTPASSMVGSSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1074


RESULT 7
US-09-801-368-108
; Sequence 108, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
;  APPLICANT: Busby, Robert
;  APPLICANT:  Cali, Brian
;  APPLICANT:  Hecht, Peter
;  APPLICANT:  Holtzman, Doug
;  APPLICANT:  Madden, Kevin
;  APPLICANT:  Maxon, Mary
;  APPLICANT:  Milne, Todd
;  APPLICANT:  No. US20020128250Alman, Thea
;  APPLICANT:  Royer, John
;  APPLICANT:  Salama, Sofie
;  APPLICANT:  Sherman, Amir
;  APPLICANT:  Silva, Jeff
;  APPLICANT:  Summers, Eric
;  TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
;  FILE REFERENCE: 109272.147
;  CURRENT APPLICATION NUMBER: US/09/801,368
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: US 09/487,558
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: US 60/160,587
;  PRIOR FILING DATE: 1999-10-20
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
;   LENGTH: 1367
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108

  Query Match             10.4%;  Score 675.5;  DB 10;  Length 1367;
  Best Local Similarity   26.2%;  Pred. No. 4.8e-25;
  Matches  378;  Conservative  177;  Mismatches  535;  Indels  353;  Gaps   65;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              | : | | :| ||:            |||              |:||     | | :|||
Db          1 MQRPFLLAYLVLSLL-----------FNSAL-----------GFPTALVPRG-SSEGTS 36

Qy         61 AN------PGDTFTLNM--PCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTC 112
               |      |   | :|    : :||   |||    |   |       |:|
Db         37 CNSIVNGCPNLDFNWHMDQQNIMQYTLDVTSVSWVQDNT-YQITIHVKGKE--------- 86

Qy        113 TVNDALKS--SIKAFGTVTLPIAFNVGGTGSSTDLEDSKC-FTAGTNTVTFN-DGDKDI- 167
                |  ||   |:|  |         : |   :| | |:   |||     ||     :|:
Db         87 --NIDLKYLWSLKIIGVTGPKGTVQLYGYNENTYLIDNPTDFTA-----TFEVYATQDVN 139

Qy        168 -------SIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSN 220
                    : :::||        ||  |             | |  : :  ||  |  |
Db        140 SCQVWMPNFQIQFE--------YLQGS-------------AAQYASSWQWGTTSFDLST 177

Qy        221 GDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYI 280
              |     |:|      :| :  ::|     | :   | :|
Db        178 G-----CNNYD---NQGHSQTDFP----GFYWNIDCDNN-------------------C 205

Qy        281 SATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA 340
                |  : | : :   |   | :|   |       :| : |:      |: |   ||::
Db        206 GGTKSSTTTSSTSESSTTTSSTSESSTTT-------SSTSESS----TTTSSTSESSTSS 254

Qy        341 VTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTT 400
```

```
                 || |  |:      |    :  ||   |||         ||  |           |:| |
Db        255 STTAPATPT-----TTSCTKEKPTPTTTS--------CTKEKP---------TPPHHDT 292

Qy        401 TTVTSEWTGTITTTTTRTN---PTDSIDTV-VVQVPLPNPTVSTTEYWSQ--SFATTTTV 454
              |  | : | |  | | :|    || |  |       |:| |: |||| |   : :|| :
Db        293 TPCTKKKTTTSKTCTKKTTTPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESS 352

Qy        455 TAP---PGGTDTVIIREPPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPPNPTVT 511
              :||   |  : |     |   : | :     : :|| : :||            | |: :
Db        353 SAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAP-----------VPTPSSS 401

Qy        512 TTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTT-VTAPPGGTD 570
              |||  |    ::|| ::    | |          | :||| |    ::||  :: | |
Db        402 TTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVPTP 461

Qy        571 SVIIREPPNHTVT--TTEYWS-----QSYATTTTVTAPPGGTDTVIIREP-PNHTVTTTE 622
              |    |  :  ||  |||  |       | :|| : :||   : |     | |  : :||
Db        462 SSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTE 521

Qy        623 YWS-----QSFATTTTVTGPPSGTDTVIIREP-PNPTVTTTEYWSQSYATTTTITAPPGE 676
                |      | :|| : : | : :  |      | | |: :|||      :::| :|:   |
Db        522 SSSAPAPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTE---:--SSSTPVTSSTTE 575

Qy        677 TDTVLIREPPNHTVTTTEYWSQSYAT----TTTVTAPPGETDTVLIREPPNHTVT--TTE 730
              : :  :   |  : :|||  |     |    ||  :: |  | :    |  :  ||  |||
Db        576 SSSAPV---PTPSSSTTESSSAPVPTPSSSTTESSSAPAPTPSSSTTESSSAPVTSSTTE 632

Qy        731 YWS-----QSYATTTTVTAP---PGGTDTVIIREP-PNPTVTTTEYWSQSFATTTTVTAP 781
                |      | :|| : :||   |  : |      | | |: :|||  | |   |:: |
Db        633 SSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTE--SSSAPVTSSTTES 690

Qy        782 PGGTDTVIIYESMSS-----SKISTSSNDITSIIPSFSR--------PHYVNSTTSD--- 825
                  |    || |:     |  :| |:      || |         |   :|||
Db        691 SSAPVTSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSA 750

Qy        826 ---LSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTN 882
                 || ||||   ||  ||     ||  : | | :|||   :     |  :: | :
Db        751 PVTSSTTESSSAPVPTP-SSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSVAPV 809

Qy        883 PDSNESSIVTSTVPTASTMSDSL-SSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNP 941
              |  : || :||: |:::  | |  ||:  :   ||      |   |:: |: :::   | |
Db        810 PTPSSSSNITSSAPSSTPFSSSTESSSVPVPTPSSSTTESSSAPVSSSTTESSVAPVPTP 869

Qy        942 LSSSVTSLTQLSSIPSVSESES-----KVTFTSNGDNQSGTHDSQSTSTEIEIV------ 990
              |||  : :  ||||  | :||      || :|:    | |  | |::|| ||
Db        870 SSSSNITSSAPSSIPFSSTTESFSTGTTVTPSSSKYPGSQTETSVSSTTETTIVPTKTTT 929

Qy        991 ---TTSSTKVLPPVVSSNTDLTSEPTN-------TREQPTTLST--TSNSITEDITTSQP 1038
                | |:| :  :| |: |:   | |:       |   ||| :|  |::| |   ||
Db        930 SVTTPSTTTITTTVCSTGTNSAGETTSGCSPKTVTTTVPTTTTTSVTTSSTTTITTTVCS 989

Qy       1039 TGDN--GDNTSSTNPVPTVATSTLASASEEDNKSGSHESA-----------------ST 1078
              || |  |: ||  :|  |: |:   | |  :  | |  ::                 ||
Db        990 TGTNSAGETTSGCSP-KTITTTVPCSTSPSETASESTTTSPTTPVTTVVSTTVVTTEYST 1048

Qy       1079 SLKPSMGENSGLTTS--TEIEATTTSPTEAPSPAV---------------SSGTD---- 1116
              | ||  ||   :||:  |:   ||.  | ||:|:|                |:||:
Db       1049 STKPG-GE---ITTTFVTKNIPTTYLTTIAPTPSVTTVTNFTPTTITTTVCSTGTNSAGE 1104

Qy       1117 ----------VTTEPTDT-----REQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLT 1161
                       || |  |       : ||| ||: | : :| |  :|  |     :|  |
Db       1105 TTSGCSPKTVTTTVPCSTGTGEYTTEATTLVTTAVTTT-VVTTESSTGTNSAGKTTTGYT 1163

Qy       1162 SSLTTGTSASTSANSELVTSG------SVTGGAVASASNDQSHSTSVTNSNSIVSN---- 1211
              :    |  :| | |  ||         ::|    ::|:|    :||| :: :|||:
Db       1164 TKSVPTTYVTTLAPSAPVTPATNAVPTTITTTECSAATNAAGETTSVCSAKTIVSSASAG 1223

Qy       1212 --------TPQTT---LSQQVTSSSPSTN----------------TFIASTYDGSGSIIQ 1244
                     || ||    :  | ||  ||                 |:| :   ||
Db       1224 ENTAPSATTPVTTAIPTTVITTESSVGTNSAGETTTGYTTKSIPTTYITTLIPGSNGAKN 1283

Qy       1245 HST 1247
              : |
```

```
Db       1284 YET 1286


RESULT 8
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12713
;   LENGTH: 2344
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12713

  Query Match             9.9%;  Score 640.5;  DB 10;  Length 2344;
  Best Local Similarity   22.2%;  Pred. No. 4.2e-23;
  Matches  323;  Conservative  269;  Mismatches  542;  Indels  323;  Gaps   46;

Qy         17 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 76
              | |:    ||  |  :|   |   |  |:      :|::      :||
Db        297 AITLGTRIDSNKSFHFSGKVNLGNKYEGHGNGGDGIGFAF-----SPG------------ 339

Qy         77 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIK--------AFGTV 128
                :| ::  | |:   :  |  | :  |:   | | | : |:           |||
Db        340 -VLGETGLNGAAVGIGGLSNAF--GFKLDTYHN-TSTPNSSAKAKADPSNVAGGGAFGAF 395

Qy        129 TLPIAFNVGGTGSSTDLED--SKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYA 186
               :: |  | :|:   |  :|     ||  || |       |: :   |
Db        396 VTTDSYGVASTYTSSSTADNAAKLNVQPTNN-TFQD------FDINYNGDT--------- 439

Qy        187 SRVMPSLNKVTTLFVAPQ----------CENGYTSGTMGFSSSNGDVAIDCSNIHIGITK 236
                  || |:  | |            ::| |: ::  ::| |  | :   :  |
Db        440 --------KVMTVKYAGQTWTRNISDWIAKSGTTNFSLSMTASTGG-ATNLQQVQFG--- 487

Qy        237 GLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAG--------YRPFIDAYISATDVNQY 288
                    :| ||::  :   |::| :|  |         |   :|  ::  :
Db        488 ------------TFEYTESAVT---QVRYVDVTTGKDIIPPKTYSGNVDQVVTIDNQQSA 532

Qy        289 TLAYTNDYTCAGSRLQS--------------KPFTLRWT---------GYKNSDAGS-- 322
               |   :||   |  |                :  | :|      ' | :  : |
Db        533 LTAKGYNYTSVDSSYASTYNDTNKTVKMTNAGQSVTYYFTDVKAPTVTVGNQTIEVGKTM 592

Qy        323 NGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPI----------------PT 364
              | ||:  |       |  || ||   | |      |  |                  |

Db        593 NPIVLTTTDNGTGTVTNTVTGLPSGLSYDSATNSIIGTPTKIGQSTVTVVSTDQANNKST 652

Qy        365 TTITTSYVGVTTSYLTKTAPIGETATVI------------------------------ 392
              || | : |  |   :|   |||: :: :
Db        653 TTFTINVVDTTAPTVT---PIGDKSSEVFSPISPINIATQDNSGNAVTNTVTGLPSGLTF 709

Qy        393 -----------VDVPYHTTTTVTSEWTGTITTTT-----TRTNPTDSIDTVVVQVPLPNP 436
                        ::   | | |::: :|  ||||     || : :||: |       :
Db        710 DSTNNTISGTPTNIGTSTITIVSTDASGNKTTTTFKYEVTRNSMSDSVST--SGSTQQSQ 767

Qy        437 TVSTTEYWSQSFATT------TTVTAPPGGTDTVIIRE--PPNHTVTTTEYWSQSFATTT 488
              :|||::  ||| :|:       |: :|    : :| : :    : :::|:|  | | :|:|
Db        768 SVSTSKADSQSASTSTSGSIMTSTSASTSKSTSVSLSDSVSASKSLSTSESNSVSSSTST 827

Qy        489 TVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTT 548
              ::       : |  :    : : : ::: | | :|  : :     :||:      : :|:
Db        828 SLVNSQSVSSS--MSGSVSKSTSLSDFISNSSSTEKSESVSTSTSDSLRTSTSLSDSVSM 885

Qy        549 TEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAP------- 601
              :   | | : : : :     : |  : :  :::::|:|  |:| :|: :::
Db        886 STSGSLSKSQSLSTSTSDSASTSQSVSDSTSNSISTSESLSESGSTSESISISNSISNSV 945

Qy        602 -------PGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTV 654
                    : :: :    : :::|:|  | | :|: :|:|  |   :  :    : ::
Db        946 SASTSKLESQSTSISLSTSDSKSMSTSESLSDSTSTSDSVSGSLSVAGSQSVSTSTSDSM 1005

Qy        655 TTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATT-------TTVT 707
              :|:|  | | :|: :: |   :: :|      : : :|:|  | | :|:       :  |
Db       1006 STSEMISDSMSTSGSLAASDSKSMSVSSSMSTSQSGSTSESLSDSISTSDSDSKSLSLST 1065

Qy        708 APPGETDTVL-----IREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTV 762
              :  | | |       :|   : : : :   ||| :|: : :     :|:       : :
Db       1066 SQSGSTSTSTSTSSSVRMSESQSTSGSMSTSQSDSTSISTSFSDSTSDS------KSAST 1119

Qy        763 TTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNST 822
              ::|  ||| :|:|:      |   |      : :| : ||| :| ||:  | |     :::
Db       1120 ASSESISQSVSTSTS-----GSVSTSTSLSTSNSERTSTSMSDSTSLSTSESDSTSDSTS 1174

Qy        823 TSDLSTFESSSMNTPTSIS-------SDGMLLSSTTLVTE------SETTTELICSDGKE 869
              ||| |  |: | :  ||||       ||     |::  ::|      ||:|:| :     :
Db       1175 TSD-SISEAISGSESTSISLSESNSTSDSESKSASAFLSESLSESTSESTSESLSGSTSD 1233

Qy        870 CSRLS-------SSSGIVTNPDSNESSIVTSTVPTAST--------------MSDSLSS 907
               : ||       |:|  ::|  |  :|| ||| :|||                | |||
Db       1234 STSLSDSNSESGSTSTSLSNSTSGSASISTSTSGSASTSTVKSESVSTSLSTSTSTSLSD 1293

Qy        908 TDGISATSSDNVSKS-----GVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESE 962
              :  :| : ||: | |       |::|  |::| ::     |:|::  |  |   | | ||
Db       1294 STSLSTSLSDSTSGSKSNSLSASMSTSDSISTRKSESLSASTSLSGSTSESESGSTSSSE 1353

Qy        963 SKVTFTS---------NGDNQSGTHD--SQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSE 1011
              ||   ||          :|     | :  | |||| : :  : :   :    :| :  ||
Db       1354 SKSDSTSMSLSMSQSTSGSTSVSTSESLSDSTSTSLSLSASMNQSGVDSNSASQSASTST 1413

Qy       1012 PTNTRE--QPTTLSTTSNSITEDITTSQPT--GDNGDNTSSTNPVPTVATSTLASASEED 1067
              |:| |    :| | || | ::  :|| |   |:   : ||:   : :||   | || :
Db       1414 STSTSESDSQSTSSYTSQSTSQSESTSTSTSLSDSTSISKSTSQSGSTSTSASLSGSESE 1473

Qy       1068 NKSGS---------HESASTSLKPSMG-ENSGLTTSTEIEATTTSPTEAPSPAVSSGTDV 1117
              : | |          || |||| |    |||  ::: : : : | :|: | : |  :|
Db       1474 SDSQSISTSTSESKSESTSTSLSDSTSTSNSGSASTSTLLSNSASASESDSSSTSL-SDS 1532

Qy       1118 TTEPTDTREQPTTLSTTSKTNSELVAT------------TQATNENGGKSPSTDLTSSLT 1165
              |:    : |  :  ::|| :||:  :|             :::|:|:|  | ||  : |.:
Db       1533 TSASMQSSESDSQSTSTSLSNSQSTSTSIRMSTIASESVSESTSESGSTSESTSESDSTS 1592

Qy       1166 TGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSS 1225
              |  | | | :      | | :| | |  | |  | ||| : | |: :    :||  | | |
Db       1593 TSLSDSQSTS----RSTSASGSASTSTSTSDSRSTSASTSTSMRT----STLDSQSMSLS 1644

Qy       1226 PSTNTFIASTYDGSGSI 1242
                ||:| :: :   | |:
Db       1645 TSTSTSVSDSTSLSDSV 1661
```

```
RESULT 9
US-10-025-380-1065
; Sequence 1065, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole L.
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  APPLICANT:  Skeiky, Yasir A. W.
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick Thomas S.
;  APPLICANT:  Carter, Darrick
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C14
;  CURRENT APPLICATION NUMBER: US/10/025,380
;  CURRENT FILING DATE:  2001-12-19
;  NUMBER OF SEQ ID NOS: 1129
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1065
;   LENGTH: 957
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-025-380-1065

  Query Match             8.4%;  Score 545.5;  DB 9;  Length 957;
  Best Local Similarity   25.9%;  Pred. No. 6e-19;
  Matches  266;  Conservative  139;  Mismatches  406;  Indels  215;  Gaps   43;

Qy        338 TTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPY 397
              |||      | | :  |       | :| ||   | ||:     : :|  :|     |
Db          6 TTAFPGSTTMPGVSQESTAS--HSSPGSTDTTLSPGSTTA-----SSLGPESTTFHSGPG 58

Qy        398 HTTTTVTSEWT---GTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTT--------EYWSQ 446
               | ||:  : |   | :  :|   : | |  | :      :|  |||         : |
Db         59 STETTLLPDNTTASGLLEASTPVHSSTGSPHTTL------SPAGSTTRQGESTTFQSWPN 112

Qy        447 SFATTTTVTAPPGGTDTVIIREPPNH---TVTTTEYWSQSFAT------TTTVTAPPGGT 497
              |  ||    |||  |   :     :|   : | | ::| |  |      :||| : |  |
Db        113 SKDTT---PAPPTTTSAFVELSTTSHGSPSSTPTTHFSASSTTLGRSEESTTVHSSPVAT 169

Qy        498 DSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYA 557
               :      |:|  :||   |     :||  : || | ::   |    :    |  :   :
Db        170 ATT-----PSPARSTT---SGLVEESTTYHSSPGSTQTMHFPESDTTSGRGEESTTSHSS 221

Qy        558 TTTTVTAPPGGTDSVIIREP------PNHTVTT-----------TEYWSQSYA---TTTT 597
              || |::: |  | | :: ||      |  | ||           :|  : |::    | |
Db        222 TTHTISSAP-STTSALVEEPTSYHSSPGSTATTHFPDSSTTSGRSEESTASHSNQDATGT 280

Qy        598 VTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATT----TTVTGPPSGTDTVIIREPPNPT 653
               :  |   | :|:: |      ||:   | |  ||    :| |   |   |     | :|
Db        281 IVLPARSTTSVLLGES-----TTSPISSGSMETTALPGSTTTPGLSEKSTTFHSSPRSPA 335

Qy        654 VTTTEYWSQSYAT---TTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPP 710
               | :   : |      :||  : || | |    :      |||   |:    :||  : |
Db        336 TTLSPASTTSSGVSEESTTSHSRPGSTHTTAFPDS-----TTTPGLSRH---STTSHSSP 387

Qy        711 GETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQ 770
              | ||| |:  | :   |||   ||    :||  : || ||| :     :|  ||   : |
Db        388 GSTDTTLL--PAS---TTTSGPSQE---STTSHSSPGSTDTAL-----SPGSTTALSFGQ 434

Qy        771 SFATTTTVTAPPGGTDTVIIYESMSSSKI--------STSSNDITSIIPSFSRPHYVNST 822
```

```
                 :||  : || | | :  :| :|| |        |:: :  |:: |: |    :
Db        435 E---STTFHSSPGSTHTTLFPDSTTSSGIVEASTRVHSSTGSPRTTLSPASSTSPGLQGE 491

Qy        823 TSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTN 882
               ::   |  :|:  ||::          ||      |:||          :   :||   |
Db        492 STAFQTHPASTHTTPST--------PSTATAPVEESTTYHRSPSSTPTTHFPASS---TT 540

Qy        883 PDSNESSIVTSTVPTAS--TMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPN 940
               :| | :  : | ||  | | : |:| |   :::  :|     :||     |   |||
Db        541 SGHSEKSTIFHSSPDASGTTPSSAHSTTSGRGESTTSRISPGSTEITTLPGST---TTPG 597

Qy        941 PLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPP 1000
               || : |:       |: : | .: :|    |: :| |  ||  ||   :   |:|    |
Db        598 -LSEASTTFYSSPRSPTTTLSPASMTSLGVGE-ESTTSRSQPGSTHSTVSPASTT---TP 652

Qy       1001 VVS--SNTDLTSEPTNTR-------------EQPTTL-----STTSNSITEDITTS--- 1036
               :|  | |  :| | :|               |:|||      || :   ||| |||
Db        653 GLSEESTTVYSSSPGSTETTVFPRSTTTSVRGEEPTTFHSRPASTHTTLFTEDSTTSGLT 712

Qy       1037 -QPTGDNGDNTSSTNPVPTVATSTLASASEEDNK-SGSHESASTSLKPSMGENSGL---- 1090
               : |   |   |:   :|  || | |   ||      |  |  |:| |:    |||
Db        713 EESTAFPGSPASTQTGLP--ATLTTADLGEESTTFPSSSGSTGTTLSPARSTTSGLVGES 770

Qy       1091 ------------------TTSTEIEATTT------SPTEAPSPAVSSGTDVTTEPTDTR 1125
                               || :  | :||      ||    ||| :: : |: | : :
Db        771 TPSRLSPSSTETTTLPGSPTTPSLSEKSTTFYTSPRSPDATLSPATTTSSGVSEESSTSH 830

Qy       1126 EQPTTLSTTS-------------KTNSELVATTQATNENGGKSPSTDLTSSL------- 1164
               || :  ||:             ||:     :|:||      || :   |||
Db        831 SQPGSTHTTAFPDSTTTSGLSQEPKTSHSSQGSTEAT-----LSPGSTTASSLGQQSTTF 885

Qy       1165 --TTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVT 1222
                 : | : :|    : :||| |       :|    |:|    |::      ::|  |
Db        886 HSSPGDTETTLLPDDTITSGLVEASTPTHSSTGSLHTTLTPASSTSAGLQEESTTFQSWP 945

Qy       1223 SSSPST 1228
              ||| :|
Db        946 SSSDTT 951
```

```
RESULT 10
US-09-922-217-1065
; Sequence 1065, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole Lynn
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C13
;  CURRENT APPLICATION NUMBER: US/09/922,217
;  CURRENT FILING DATE:  2001-08-03
;  NUMBER OF SEQ ID NOS: 1124
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1065
;   LENGTH: 957
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-922-217-1065

  Query Match             8.4%;  Score 545.5;  DB 10;  Length 957;
  Best Local Similarity   25.9%;  Pred. No. 6e-19;
  Matches  266;  Conservative  139;  Mismatches  406;  Indels  215;  Gaps   43;
```

```
Qy       338 TTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPY 397
Db         6 TTAFPGSTTMPGVSQESTAS--HSSPGSTDTTLSPGSTTA-----SSLGPESTTFHSGPG 58

Qy       398 HTTTTVTSEWT---GTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTT--------EYWSQ 446
Db        59 STETTLLPDNTTASGLLEASTPVHSSTGSPHTTL------SPAGSTTRQGESTTFQSWPN 112

Qy       447 SFATTTTVTAPPGGTDTVIIREPPNH---TVTTTEYWSQSFAT------TTTVTAPPGGT 497
Db       113 SKDTT---PAPPTTTSAFVELSTTSHGSPSSTPTTHFSASSTTLGRSEESTTVHSSPVAT 169

Qy       498 DSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYA 557
Db       170 ATT-----PSPARSTT---SGLVEESTTYHSSPGSTQTMHFPESDTTSGRGEESTTSHSS 221

Qy       558 TTTTVTAPPGGTDSVIIREP------PNHTVTT-----------TEYWSQSYA---TTTT 597
Db       222 TTHTISSAP-STTSALVEEPTSYHSSPGSTATTHFPDSSTTSGRSEESTASHSNQDATGT 280

Qy       598 VTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATT----TTVTGPPSGTDTVIIREPPNPT 653
Db       281 IVLPARSTTSVLLGES-----TTSPISSGSMETTALPGSTTTPGLSEKSTTFHSSPRSPA 335

Qy       654 VTTTEYWSQSYAT---TTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPP 710
Db       336 TTLSPASTTSSGVSEESTTSHSRPGSTHTTAFPDS-----TTTPGLSRH---STTSHSSP 387

Qy       711 GETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQ 770
Db       388 GSTDTTLL--PAS---TTTSGPSQE---STTSHSSPGSTDTAL-----SPGSTTALSFGQ 434

Qy       771 SFATTTTVTAPPGGTDTVIIYESMSSSKI--------STSSNDITSIIPSFSRPHYVNST 822
Db       435 E---STTFHSSPGSTHTTLFPDSTTSSGIVEASTRVHSSTGSPRTTLSPASSTSPGLQGE 491

Qy       823 TSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTN 882
Db       492 STAFQTHPASTHTTPST--------PSTATAPVEESTTYHRSPSSTPTTHFPASS---TT 540

Qy       883 PDSNESSIVTSTVPTAS--TMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPN 940
Db       541 SGHSEKSTIFHSSPDASGTTPSSAHSTTSGRGESTTSRISPGSTEITTLPGST---TTPG 597

Qy       941 PLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPP 1000
Db       598 -LSEASTTFYSSPRSPTTTLSPASMTSLGVGE-ESTTSRSQPGSTHSTVSPASTT---TP 652

Qy      1001 VVS--SNTDLTSEPTNTR--------------EQPTTL-----STTSNSITEDITTS--- 1036
Db       653 GLSEESTTVYSSSPGSTETTVFPRSTTTSVRGEEPTTFHSRPASTHTTLFTEDSTTSGLT 712

Qy      1037 -QPTGDNGDNTSSTNPVPTVATSTLASASEEDNK-SGSHESASTSLKPSMGENSGL---- 1090
Db       713 EESTAFPGSPASTQTGLP--ATLTTADLGEESTTFPSSSGSTGTTLSPARSTTSGLVGES 770

Qy      1091 ------------------TTSTEIEATTT------SPTEAPSPAVSSGTDVTTEPTDTR 1125
Db       771 TPSRLSPSSTETTTLPGSPTTPSLSEKSTTFYTSPRSPDATLSPATTTSSGVSEESSTSH 830

Qy      1126 EQPTTLSTTS--------------KTNSELVATTQATNENGGKSPSTDLTSSL------- 1164
Db       831 SQPGSTHTTAFPDSTTTSGLSQEPKTSHSSQGSTEAT-----LSPGSTTASSLGQQSTTF 885

Qy      1165 --TTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVT 1222
Db       886 HSSPGDTETTLLPDDTITSGLVEASTPTHSSTGSLHTTLTPASSTSAGLQEESTTFQSWP 945

Qy      1223 SSSPST 1228
Db       946 SSSDTT 951
```

```
RESULT 11
US-09-833-263-1065
; Sequence 1065, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Meagher, Madeleine J.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C12
;  CURRENT APPLICATION NUMBER: US/09/833,263
;  CURRENT FILING DATE:  2001-04-10
;  NUMBER OF SEQ ID NOS: 1093
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1065
;   LENGTH: 957
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-833-263-1065

  Query Match             8.4%;  Score 545.5;  DB 10;  Length 957;
  Best Local Similarity   25.9%;  Pred. No. 6e-19;
  Matches  266;  Conservative  139;  Mismatches  406;  Indels  215;  Gaps   43;

Qy       338 TTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPY 397
Db         6 TTAFPGSTTMPGVSQESTAS--HSSPGSTDTTLSPGSTTA-----SSLGPESTTFHSGPG 58

Qy       398 HTTTTVTSEWT---GTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTT--------EYWSQ 446
Db        59 STETTLLPDNTTASGLLEASTPVHSSTGSPHTTL------SPAGSTTRQGESTTFQSWPN 112

Qy       447 SFATTTTVTAPPGGTDTVIIREPPNH---TVTTTEYWSQSFAT------TTTVTAPPGGT 497
Db       113 SKDTT---PAPPTTTSAFVELSTTSHGSPSSTPTTHFSASSTTLGRSEESTTVHSSPVAT 169

Qy       498 DSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYA 557
Db       170 ATT-----PSPARSTT---SGLVEESTTYHSSPGSTQTMHFPESDTTSGRGEESTTSHSS 221

Qy       558 TTTTVTAPPGGTDSVIIREP------PNHTVTT-----------TEYWSQSYA---TTTT 597
Db       222 TTHTISSAP-STTSALVEEPTSYHSSPGSTATTHFPDSSTTSGRSEESTASHSNQDATGT 280

Qy       598 VTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATT----TTVTGPPSGTDTVIIREPPNPT 653
Db       281 IVLPARSTTSVLLGES-----TTSPISSGSMETTALPGSTTTPGLSEKSTTFHSSPRSPA 335

Qy       654 VTTTEYWSQSYAT---TTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPP 710
Db       336 TTLSPASTTSSGVSEESTTSHSRPGSTHTTAFPDS-----TTTPGLSRH---STTSHSSP 387

Qy       711 GETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQ 770
Db       388 GSTDTTLL--PAS---TTTSGPSQE---STTSHSSPGSTDTAL-----SPGSTTALSFGQ 434

Qy       771 SFATTTTVTAPPGGTDTVIIYESMSSSKI--------STSSNDITSIIPSFSRPHYVNST 822
Db       435 E---STTFHSSPGSTHTTLFPDSTTSSGIVEASTRVHSSTGSPRTTLSPASSTSPGLQGE 491

Qy       823 TSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTN 882
Db       492 STAFQTHPASTHTTPST--------PSTATAPVEESTTYHRSPSSTPTTHFPASS---TT 540

Qy       883 PDSNESSIVTSTVPTAS--TMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPN 940
Db       541 SGHSEKSTIFHSSPDASGTTPSSAHSTTSGRGESTTSRISPGSTEITTLPGST---TTPG 597

Qy       941 PLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPP 1000
```

```
Db        598 -LSEASTTFYSSPRSPTTTLSPASMTSLGVGE-ESTTSRSQPGSTHSTVSPASTT---TP 652

Qy       1001 VVS--SNTDLTSEPTNTR--------------EQPTTL-----STTSNSITEDITTS--- 1036
Db        653 GLSEESTTVYSSSPGSTETTVFPRSTTTSVRGEEPTTFHSRPASTHTTLFTEDSTTSGLT 712

Qy       1037 -QPTGDNGDNTSSTNPVPTVATSTLASASEEDNK-SGSHESASTSLKPSMGENSGL---- 1090
Db        713 EESTAFPGSPASTQTGLP--ATLTTADLGEESTTFPSSSGSTGTTLSPARSTTSGLVGES 770

Qy       1091 ------------------TTSTEIEATTT------SPTEAPSPAVSSGTDVTTEPTDTR 1125
Db        771 TPSRLSPSSTETTTLPGSPTTPSLSEKSTTFYTSPRSPDATLSPATTTSSGVSEESSTSH 830

Qy       1126 EQPTTLSTTS--------------KTNSELVATTQATNENGGKSPSTDLTSSL------- 1164
Db        831 SQPGSTHTTAFPDSTTTSGLSQEPKTSHSSQGSTEAT-----LSPGSTTASSLGQQSTTF 885

Qy       1165 --TTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVT 1222
Db        886 HSSPGDTETTLLPDDTITSGLVEASTPTHSSTGSLHTTLTPASSTSAGLQEESTTFQSWP 945

Qy       1223 SSSPST 1228
Db        946 SSSDTT 951


RESULT 12
US-10-142-515-11
; Sequence 11, Application US/10142515
; Publication No. US20030078399A1
; GENERAL INFORMATION:
;  APPLICANT: SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH
;  APPLICANT:  Lloyd, Kenneth O.
;  APPLICANT:  Yin, Beatrice W.T.
;  TITLE OF INVENTION: Nucleic Acid Sequence Encoding Ovarian Antigen, CA125, and Uses T
;  FILE REFERENCE: 649-A-US
;  CURRENT APPLICATION NUMBER: US/10/142,515
;  CURRENT FILING DATE:  2002-07-23
;  PRIOR APPLICATION NUMBER: US 60/290,480
;  PRIOR FILING DATE: 2001-05-11
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
;   LENGTH: 5877
;   TYPE: PRT
;   ORGANISM: Human Being
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (1)..(5877)
;   OTHER INFORMATION: Amino acid sequence of MUC16B
US-10-142-515-11

  Query Match             8.0%;  Score 522.5;  DB 9;  Length 5877;
  Best Local Similarity   22.5%;  Pred. No. 5.5e-17;
  Matches  336;  Conservative  211;  Mismatches  568;  Indels  379;  Gaps   65;

Qy         58 GTSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDA 117
Db       1861 GITSFPESRFTMSV----TESTHHLSTDLLP----------SAETIS-----TGTVMPS 1900

Qy        118 LKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKST 177
Db       1901 LSEAMTSFATTGVPRA--ISGSGSPFSRTESGPGDATLSTIA---ESLPSSTPVPFSSST 1955

Qy        178 VDPSAYLYASRVMPSLNKVTTLFVAPQ-------CENGYTSG--TMGFSSSNGDVAIDCS 228
Db       1956 FTTT----DSSTIPALHEITSSSATPYRVDTSLGTESSTTEGRLVMGTESST-------T 2004

Qy        229 NIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYR-PFIDAYISATD-VN 286
Db       2005 EGRLVMVSTLDTSSQPGRTSSSPILDTRMTESVEL--GTVTSAYQVPSLSTRLTRTDGIM 2062

Qy        287 QYTLAYTNDYTCAG----------SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTD 336
Db       2063 EHITKIPNEAAHRGTIRPVKGPQTSTSPASPKGLHTGGTKRMETTTTALKTTTTALKTTS 2122

Qy        337 STTAVTTL---------PFNPSVDKTKTIE---------ILQPIPTTT--ITTSYVGVTT 376
Db       2123 RATLTTSVYTPTLGTLTPLNASMQMASTIPTEMMITTPYVFPDVPETTSSLATSLGAETS 2182

Qy        377 SYLTKTAPIGETATVIVDVPYHTTTTVTS----EWTGTITTTTTRTNPTDSIDTVVVQVP 432
Db       2183 TALPRTTP------SVFNRESETTASLVSRSGAERSPVIQTLDVSSSEPDTTASWVIHPA 2236

Qy        433 LPNPTVS-TTEYWSQSFATTTTVTAPPGGTD----------------TVIIREPPNHTV 474
Db       2237 ETIPTVSKTTPNFFHSELDTVSSTATSHGADVSSAIPTNISPSELDALTPLVTISGTDTS 2296

Qy        475 TTTEYWSQS---FATTTTVTAPPGGTDSVIIREPPN---------PTVTTTEYWSQSFAT 522
Db       2297 TTFPTLTKSPHETETRTTWLTHPAETSSTIPRTIPNFSHHESDATPSIATSP-GAETSSA 2355

Qy        523 TTTVTAPPGGTDSVII--------REPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVII 574
Db       2356 IPIMTVSPGAEDLVTSQVTSSGTDRNMTIPTLTLSP--GEPKTIASLVTHPEAQTSSAI- 2412

Qy        575 REPPNHTVT-------TTEYWSQSYATTTT---VTAPPGGTDTVI------IREPPNHTV 618
Db       2413 ---PTSTISPAVSRLVTSMVTSLAAKTSTTNRALTNSPGEPATTVSLVTHSAQTSPTVPW 2469

Qy        619 TTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTT------TEYWSQSYATTTT--- 669
Db       2470 TTSIFFHSKSDTTPSMTTSHGAESSSAV---PTPTVSTEVPGVVTPLVTSSRAVISTTIP 2526

Qy        670 -ITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTT 728
Db       2527 ILTLSPGEPETT-----PS---MATSHGEEASSAIPTPTVSPGVPGVVTSLVTSSRAVTS 2578

Qy        729 TEY----WSQSYATTTTVTAPPGGTDT-----VIIREPPNPTVTTTEYWSQSFATTT--T 777
Db       2579 TTIPILTFSLGEPETTPSMATSHGTEAGSAVPTVLPEVPG-MVTSLVASSRAVTSTTLPT 2637

Qy        778 VTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSM--- 834
Db       2638 LTLSPGEPETT---PSMATSHGAEASSTVPTVSPEV--PGVVTSLVTSSSGVNSTSIPTL 2692

Qy        835 --------NTPTSISSDGMLLSS---------------TTLVT---------------- 854
Db       2693 ILSPGELETTPSMATSHGAEASSAVPTPTVSPGVSGVVTPLVTSSRAVTSTTIPILTLSS 2752

Qy        855 -ESETTTELICSDGKECSR--LSSS---SGIVTNPDSNESSIVTSTVPTASTMSDSLSST 908
Db       2753 SEPETTPSMATSHGVEASSAVLTVSPEVPGMVTFLVTSSRAVTSTTIPTLTISSDEPETT 2812

Qy        909 DGISATSSDNVSKS----GVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSV------ 958
Db       2813 TSLVTHSEAKMISAIPTLGVSPTVQGLVTSLVTSSGSETSAFSNLTVASSQPETIDSWVA 2872

Qy        959 ---SESESKV-TFT-SNGDNQSG----THDSQSTST----------------------E 986
Db       2873 HPGTEASSVVPTLTVSTGEPFTNISLVTHPAESSSTLPRTTSRFSHSELDTMPSTVTSPE 2932

Qy        987 IEIVTTSSTKVLPP--------VVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQP 1038
Db       2933 AESSSAISTTISPGIPGVLTSLVTSSGRDISATFPTVPESPHESEATASWVTHPAVTSTT 2992

Qy       1039 TGDNGDNTSSTNP--VPTVATSTLASA----------------SEEDNKSGSHESAS-T 1078
Db       2993 VPRTTPNYSHSEPDTTPSIATSPGAEATSDFPTITVSPDVPDMVTSQVTSSGTDTSITIP 3052

Qy       1079 SLKPSMGE---NSGLTTSTEIEATTTSPTEAPSP---------AVSSGTDVTT---EPTD 1123
Db       3053 TLTLSSGEPETTTSFITYSETHTSSAIPTLPVSPDASKMLTSLVISSGTDSTTTFPTLTE 3112

Qy       1124 TREQPTT----LSTTSKTNSELVATTQ-------------ATNENGGKSPS--------- 1157
```

```
Db       3113 TPYEPETTAIQLIHPAETNTMVPRTTPKFSHSKSDTTLPVAITSPGPEASSAVSTTTISP 3172

Qy       1158 --TDLTSSL--TTGTSASTS-----------ANSELVTSGSVTGGAVASASNDQSHSTS 1201
Db       3173 DMSDLVTSLVPSSGTDTSTTFPTLSETPYEPETTATWLTHPAETSTTVSGTIPNFSHRGS 3232

Qy       1202 VTNSNSIV-------SNTPQTTL--------SQQVTSSSPSTNTFIASTYDGSG 1240
Db       3233 DTAPSMVTSPGVDTRSGVPTTTIPPSIPGVVTSQVTSSATDTSTAIPTLTPSPG 3286


RESULT 13
US-09-801-368-106
; Sequence 106, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
;  APPLICANT: Busby, Robert
;  APPLICANT:  Cali, Brian
;  APPLICANT:  Hecht, Peter
;  APPLICANT:  Holtzman, Doug
;  APPLICANT:  Madden, Kevin
;  APPLICANT:  Maxon, Mary
;  APPLICANT:  Milne, Todd
;  APPLICANT:  No. US20020128250Alman, Thea
;  APPLICANT:  Royer, John
;  APPLICANT:  Salama, Sofie
;  APPLICANT:  Sherman, Amir
;  APPLICANT:  Silva, Jeff
;  APPLICANT:  Summers, Eric
;  TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
;  FILE REFERENCE: 109272.147
;  CURRENT APPLICATION NUMBER: US/09/801,368
;  CURRENT FILING DATE:  2001-03-07
;  PRIOR APPLICATION NUMBER: US 09/487,558
;  PRIOR FILING DATE: 2000-01-19
;  PRIOR APPLICATION NUMBER: US 60/160,587
;  PRIOR FILING DATE: 1999-10-20
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 106
;   LENGTH: 1169
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-801-368-106

  Query Match           7.7%;  Score 501;  DB 10;  Length 1169;
  Best Local Similarity  23.1%;  Pred. No. 1e-16;
  Matches  313;  Conservative  208;  Mismatches  528;  Indels  308;  Gaps   53;

Qy          6 TLLFLYLSIASAKTIT-----------GVFDSFNSLTWSNAANYAFKGPGYPTWNAVLG 53
Db         11 TGLFL-LSVANVALGTTEACLPAGEKKNGMTINFYQYSLKDSSTYS--NPSYMAYGYADA 67

Qy         54 WSLDGTSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGE-------EFTT 106
Db         68 EKLGSVSGQTKLSIDYSIPCNGASDTCACSDD---DATEYSASQVVPVKRGVKLCSDNTT 124

Qy        107 FSTLT-------CTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVT 159
Db        125 LSSKTEKRENDDCDQGAAYWSS-DLFGFYTTPTNVTVEMTG--------YFLPPKTGTYT 175

Qy        160 FNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSS 219
Db        176 FG-------------FATVDDSAIL---------------------------------SV 189

Qy        220 NGDVAIDCSNIH---IGIT----KGLNDWN--YPVSSESFSYTKTCTSNGIQIKYQN-VP 269
Db        190 GGNVAFECCKQEQPPITSTDFTINGIKPWNADAPTDIKGSTYMYAGYYYPIKIVYSNAVS 249

Qy        270 AGYRPFIDAYISATDVNQ----YTLAYTND---------------YTCAGSRLQSKPFTL 310
Db        250 WGTLPVSVVLPDGTEVNDDFEGYVFSFDDNATQAHCSVPNPAEHARTCVSSATSS----- 304

Qy        311 RWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTS 370
Db        305 -WSSSE-----------VCTECTETESTSYVTPYVTSSSWSSSEVCTECTETESTSTSTP 352

Qy        371 YVGVTTSYLTKT-APIGETATVIVDVPYHTTTTVTSEWTGTITTTT---TRTNPTDSIDT 426
Db        353 YVTSSSSSSSEVCTECTETESTSYVTPYVSSSTAAANYTSSFSSSSEVCTECTETESTST 412

Qy        427 VVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFAT 486
Db        413 STPYV--------TSSSWSSSEVCTECTE-----TESTSYVTPYVSSSTAAANYTSSFSS 459

Qy        487 TTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTV 546
Db        460 SSEVCTECTETES---------TSTSTPYVTSSSSSSSEVCTECTETESTSYVTPYVSSS 510

Qy        547 TTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTD 606
Db        511 TAAANYTSSFSSSSEVCTECTETES---------TSTSTPYVTSSSWSSSEVCTECTETE 561

Qy        607 TVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYAT 666
Db        562 STSYVTPYVSSSTAAANYTSSFSSSSEVCTECTETES---------TSTSTPYATSSTGT 612

Qy        667 TTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTV 726
Db        613 ATSFTASTSNTMTSLVQ------TDTTVSFSLSSTVSEHTNAP-----TSSVESNASTFI 661

Qy        727 TTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVT-----TTEYWSQSFATTTTVTAP 781
Db        662 SSNKGSVKSYVTSSIHSITP--------MYPSNQTVTSSSVVSTPITSESSESSASVTIL 713

Qy        782 PGGTDTVIIYESMSSSKISTSSNDITSIIPSF---SRPHYVNSTTSDLSTFESSSMNTPT 838
Db        714 P-STITSEFKPSTMKTKVVSISSSPTNLITSYDTTSKDSTVGSSTSSVSLISSISLPSSY 772

Qy        839 SISSD----GMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTST 894
Db        773 SASSEQIFHSSIVSSNGQALTSFSSTKVSSSESSESHRTSPTT-------SSESGIKSSG 825

Qy        895 VPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSS 954
Db        826 VEIESTSTSSFSFHE--TSTASTSVQISSQFVTPSSPISTVA----PRSTGLNSQTE--- 876

Qy        955 IPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTD------- 1007
Db        877 ----STNSSKETMSSE-NSASVMPSSSATSPKTGKVTSDET-------SSGFSRDRTTVY 924

Qy       1008 -LTSEPTNTREQPTTLSTT-------SNSITEDITTSQPTGDNGDNTSSTNPVPTVATST 1059
Db        925 RMTSETPSTNEQTTLITVSSCESNSCSNTVSSAVVSTATTTINGITTEYTTWCPLSATE- 983

Qy       1060 LASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAP---SPAVSSGTD 1116
Db        984 LTTVSKLESE----------------EKTTLITVTSCESGVCSETASPAIVSTATATVND 1027

Qy       1117 VTTEPTDTREQPTT-LSTTSKTNSEL--------VATTQATNENGGKSPSTDLTSSLT-T 1166
Db       1028 VVTVYSTWSPQATNKLAVSSDIENSASKASFVSEAAETKSISRNNNFVPTSGTTSIETHT 1087

Qy       1167 GTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSI-VSNTPQTTLSQQVTSSS 1225
Db       1088 TTTSNASENSDNV-----------SASEAVS-SKSVTNPVLISVSQQPRGTPASSMIGS- 1134

Qy       1226 PSTNTFIASTYDGSGSIIQHSTWLYGLITL--LSLFI 1260
Db       1135 -STASLEMSSYLG---IANH------LLTNSGISIFI 1161


RESULT 14
US-09-864-761-36047
; Sequence 36047, Application US/09864761
```

```
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36047
;   LENGTH: 688
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AB023048.1
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =   1
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =   2
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.7
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4
US-09-864-761-36047

  Query Match             7.1%;  Score 463;  DB 10;  Length 688;
  Best Local Similarity   26.5%;  Pred. No. 4e-15;
  Matches  229;  Conservative  116;  Mismatches  312;  Indels  206;  Gaps   37;

Qy        332 RTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATV 391
              ||:| :  : ||     :|  | :         ||::|   | ||  :|     |   |

Db          3 RTITSTEGSETT-----TVSATGS--------ETTVSTEGSGTTTVSIT-----GSETTK 44

Qy        392 IVDVPYHTTTTVT--SE-WTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSF 448
              :     ||||  |  ||  | :|| : | |  |:  :|          | ::||    :
Db         45 VSTTGSETTTTSTEGSEITTASITGSETTTASTEGSET----------TTASTEGSETTS 94

Qy        449 ATTT---TVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREP 505
              |:||   | ||    ::| :               :    : ||: :  |   : :
Db         95 ASTTGSETTTASTTSSETTM--------------ASIMGSETTMASTIGSETTKVSTAS 139

Qy        506 PNPTVTTTEYWSQSFATTT---TVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTV 562
                |   ||    : |:||   | |    |::::      : | |||     |   : |
Db        140 SKMTTVFTENSETTIASTTASETTTVSTAGSETIPASTAGSETTTTT-----STEGSETT 194

Qy        563 TAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTE 622
              ||   |:::       : | | |   |:      | ||   |::|       : | | :
Db        195 TASTEGSETTTASTESSETTTATTIGSE------TTTASTEGSETTTSTEGSETTTAST 248

Qy        623 YWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLI 682
                |:   || : ||  : | :    |    |  :||      :  ||::   || ||
Db        249 EGSE--ITTVSTTGSETTTAST---EGSETTTASTE-----GSELTTVSTTGSETITVSA 298

Qy        683 REPPNHTVTT----TEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYAT 738
                  ||||    |   | : : ||||:    || |  |      ||::|     :
Db        299 EGSETTTVTTMGSETTTASTAGSETTTVSTAGSETTTASIEGSETTTVSSTG------SE 352

Qy        739 TTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMS-SS 797
              ||||:    ||:|         |:|:||      : |||||    |::|  :| : | ::
Db        353 TTTVSTT--GTET---------TITSTE-----GSETTTVTT--AGSETTAVYTTGSETT 394

Qy        798 KISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPT-SISSDGMLLSSTTLVTES 856
                ||  :: |:          |::| |: :|  :: : | | | |  |  :||     |
Db        395 TTSTEGSETTT----------VSTTGSETTTASTADLETTTVSTSGSGTTTAST---AGS 441

Qy        857 ETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSS 916
              ||||  |             :|  |   | | |  |:                  :| |||
Db        442 ETTTVYI------------TGSKTTTASTEGSEATT----------------VSTTSS 471

Qy        917 DNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSG 976
              :  : |    || : :||: ||       |:  | :|:| | : : |          :
Db        472 ETTTAS----TTGSEMTTVFTT-------VSETTTVSTIGSEATTSS----------AA 509

Qy        977 THDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTS 1036
              :: :|||    ||:||       | |                  ||:::   : ||:
Db        510 GSEATTTSTEGSETTTAST------AGSET----------------TTASTAGSETTTA 546

Qy       1037 QPTGDNGDNTSSTNPVPTVA--TSTLASASEEDNKS---GSHES-AST-SLKPSMGENSG 1089
                :|      | ||   |    ||| :||  | | :   ||  : ||| ||:|:    :|
Db        547 STSG------SETNTACTTGSETSTPSSAGSETNTAFIIGSESTIASTASLEPTATSLTG 600

Qy       1090 LTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSELVATTQATN 1149
                |:|   : | |   | |  ||| | | |: ||   || | :  | ||  :       |
Db        601 SETTT--VSITASGATAASTTVSSTTFVLTKATDVSIQPITNTPMSGTNPHVFFEPTHFN 658

Qy       1150 ENGGKSPSTDLTSSLTTGTSAST 1172
                :|     |         : |||
Db        659 SSGNHQLFTCFYHLCPGSSQAST 681


RESULT 15
US-09-769-787-109
; Sequence 109, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
;  APPLICANT: Microbial Technics Limited
;  APPLICANT:  Gilbert, Christophe FG
;  APPLICANT:  Hansbro, Philip M
;  TITLE OF INVENTION: Proteins
;  FILE REFERENCE: PWC/P21129WO
;  CURRENT APPLICATION NUMBER: US/09/769,787
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: GB 9816337.1
;  PRIOR FILING DATE: 1998-03-27
```

```
;  PRIOR APPLICATION NUMBER: US 60/125164
;  PRIOR FILING DATE: 1999-03-19
;  NUMBER OF SEQ ID NOS: 388
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 109
;   LENGTH: 1236
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-769-787-109

  Query Match             6.6%;  Score 428;  DB 9;  Length 1236;
  Best Local Similarity   23.3%;  Pred. No. 3.7e-13;
  Matches  217;  Conservative  156;  Mismatches  433;  Indels  126;  Gaps   26;

Qy       327 IVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTT-TITTSYVGVTTSYLTKTAPI 385
Db        22 IVAAGAVISGTVATQTKVFTNESAVLEKTVEKTDALATNDTVVLGTISTSNSASSTSLSA 81

Qy       386 GETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWS 445
Db        82 SESA----------STSASESASTSASTSASTSASESASTSA----STSISASSTVVGS 126

Qy       446 QSFATTTTVTAPPGGTDTVIIREPP--NHTVTTTEYWSQSFATTTTVTAPPGGTDSV--I 501
Db       127 QT-AAATEATAKKVEED----RKKPASDYVASVTNVNLQSYAKRR-----KRSVDSIEQL 176

Qy       502 IREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTT 561
Db       177 LASIKNAAV---------FSGNTIVNGAPAINASLNI----------------AKSET 209

Qy       562 VTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIRE-PPNHTVTT 620
Db       210 KVYTGEGVDSV-YRVPIYYKLKVTNDGSKLTFTYTVTYVNPKTNDLGNISSMRPGYSIYN 268

Qy       621 TEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTV 680
Db       269 SGTSTQTMLTLGSDLGKPSGVKNYITDKNGRQVL--------SYNTSTMTTQGSGYT--- 317

Qy       681 LIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTT 740
Db       318 ----WGNGAQMNGFFAKKGYGLTSSWTVPITGTDTSFTFTP----------YAART 359

Qy       741 T---VTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSS 797
Db       360 DRIGINYFNGGGKVV-------ESSTTSQSLSQS--KSLSVSASQSASASA---STSASA 407

Qy       798 KISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESE 857
Db       408 SASTSASASASTSASASA-----STSASVSASTSASASASTSASA-----SASTSASESA 457

Qy       858 TTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSD 917
Db       458 STSASASASTSASASASTSASASASTSASESASTSASASASTSASESASTSASASASTSA 517

Qy       918 NVSKS-----GVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGD 972
Db       518 SASASTSASGSASTSTSASASTSASASASTSASASASISASESASTSASESASTSTSASA 577

Qy       973 NQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITED 1032
Db       578 STSAS-ESASTSASASASTSASASASTSASASASTSASASTSASES----ASTSASASAS 632

Qy      1033 ITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTT 1092
Db       633 TSASASASTSASASASTSASASASTSASVSASTSASASAS-TSASASASTSASESA--ST 689

Qy      1093 STEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSELVATTQATNENG 1152
Db       690 SASASASTSASASASTSASASASTSASESASTSASASASTSASESASTSASASASTSASA 749

Qy      1153 GKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNT 1212
Db       750 SASTSASGSASTSTSASASTSASASASTSASASASISASESASTSASESASTSTSASAST 809

Qy      1213 PQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQ 1244
Db       810 SASESASTSASASASTSASASASTSASASARQ 841

Search completed: June 11, 2003, 17:22:23
Job time : 71 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 17:11:32 ; Search time 18 Seconds
(without alignments)
2059.605 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQQFTLLFLYLSIASAKTI.....SIIQHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	940	14.5	1537	1	US-08-325-267A-2	Sequence 2, Appli
2	651	10.0	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
3	516	7.9	894	3	US-08-362-525-22	Sequence 22, Appl
4	516	7.9	894	3	US-08-971-692-15	Sequence 15, Appl
5	488	7.5	1721	3	US-08-928-361B-6	Sequence 6, Appli
6	487	7.5	1721	3	US-08-700-651-5	Sequence 5, Appli
7	487	7.5	1837	3	US-08-928-361B-5	Sequence 5, Appli
8	468.5	7.2	862	1	US-08-325-267A-4	Sequence 4, Appli
9	404.5	6.2	2035	1	US-08-046-585-5	Sequence 5, Appli
10	404.5	6.2	2035	1	US-08-393-703-5	Sequence 5, Appli
11	404.5	6.2	2035	5	PCT-US93-11721-5	Sequence 5, Appli
12	392	6.0	1638	4	US-09-071-035-258	Sequence 258, App
13	392	6.0	1638	4	US-09-071-035-262	Sequence 262, App
14	392	6.0	1638	4	US-09-071-035-266	Sequence 266, App
15	364.5	5.6	1022	4	US-07-757-022B-84	Sequence 84, Appl
16	364.5	5.6	1320	4	US-07-757-022B-46	Sequence 46, Appl
17	364.5	5.6	1320	4	US-07-757-022B-60	Sequence 60, Appl
18	364.5	5.6	1361	4	US-07-757-022B-40	Sequence 40, Appl
19	362	5.6	1049	4	US-07-757-022B-58	Sequence 58, Appl
20	362	5.6	1140	4	US-07-757-022B-104	Sequence 104, App
21	362	5.6	1313	4	US-07-757-022B-142	Sequence 142, App
22	362	5.6	1314	4	US-07-757-022B-50	Sequence 50, Appl
23	362	5.6	1354	4	US-07-757-022B-48	Sequence 48, Appl
24	362	5.6	1363	4	US-07-757-022B-52	Sequence 52, Appl
25	362	5.6	1404	4	US-07-757-022B-2	Sequence 2, Appli
26	362	5.6	1404	4	US-07-757-022B-62	Sequence 62, Appl
27	361.5	5.6	1038	4	US-07-757-022B-74	Sequence 74, Appl
28	361.5	5.6	1270	4	US-07-757-022B-44	Sequence 44, Appl
29	361.5	5.6	1311	4	US-07-757-022B-42	Sequence 42, Appl
30	359.5	5.5	941	4	US-07-757-022B-14	Sequence 14, Appl
31	350	5.4	650	3	US-08-362-525-2	Sequence 2, Appli
32	323	5.0	907	3	US-08-783-774-2	Sequence 2, Appli
33	323	5.0	907	4	US-09-328-599A-1	Sequence 1, Appli
34	323	5.0	907	5	PCT-US95-04611A-19	Sequence 19, Appl
35	322	5.0	1481	2	US-08-616-844-40	Sequence 40, Appl
36	322	5.0	1481	2	US-08-599-654-40	Sequence 40, Appl
37	322	5.0	1481	3	US-08-944-868A-40	Sequence 40, Appl
38	322	5.0	1481	3	US-08-944-423A-40	Sequence 40, Appl
39	322	5.0	1481	3	US-08-944-496-40	Sequence 40, Appl
40	310.5	4.8	249	3	US-08-700-651-15	Sequence 15, Appl
41	310.5	4.8	249	3	US-08-928-361B-20	Sequence 20, Appl
42	304.5	4.7	878	4	US-09-556-706B-2	Sequence 2, Appli
43	304.5	4.7	2409	6	5180808-2	Patent No. 5180808
44	304	4.7	1203	4	US-09-351-200-2	Sequence 2, Appli
45	303	4.7	216	3	US-08-928-361B-27	Sequence 27, Appl

ALIGNMENTS

```
RESULT 1
US-08-325-267A-2
; Sequence 2, Application US/08325267A
; Patent No. 5585271
;  GENERAL INFORMATION:
;    APPLICANT:  WATARI, JUNJI
;    APPLICANT:  TAKATA, YOSHIHIRO
;    APPLICANT:  OGAWA, MASAHIRO
;    APPLICANT:  PENTTILA, MERJA
;    APPLICANT:  ONNELA, MAIJA-LEENA
;    APPLICANT:  KERANEN, SIRKKA
;    TITLE OF INVENTION:  YEAST AGGLUTINATION GENES AND YEAST
;    TITLE OF INVENTION:  CONTAINING THEM
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;      STREET:  1755 S. JEFFERSON DAVIS HWY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,267A
;      FILING DATE:  18-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP PCT/JP94/00290
;      FILING DATE:  24-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 38871/1993
;      FILING DATE:  26-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2589-023-0XPCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 1537 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
```

US-08-325-267A-2

Query Match 14.5%; Score 940; DB 1; Length 1537;
Best Local Similarity 26.0%; Pred. No. 5.9e-46;
Matches 417; Conservative 244; Mismatches 523; Indels 418; Gaps 85;

```
Qy          2 LQQFTLLFLYLSIASAKT---------ITGVFDSF------NSLTWSNAANYAF------ 40
              |  |||| |  |:|| |           :|:  :|      :| |:|||| |:
Db         10 LAVFTLLAL-TSVASGATEACLPAGQRKSGMNINFYQYSLKDSSTYSNAAYMAYGYASKT 68

Qy         41 ----------------------------KGPGYPTW-----------NAVLGWSLD--G 58
                                          :   |  |             :  || |  |
Db         69 KLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGACSNSQGIAYWSTDLFG 128

Qy         59 TSANPGDTFTLNM----------PCVFKYTTSQ----------------------TSVD 85
                 |  :  || |          ||: |                          || :
Db        129 FYTTPTNV-TLEMTGYFLPPQTGSYTFKFATVDDSAILSVGGATAFNCCAQQQPPITSTN 187

Qy         86 LTADGVK----------YATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGT----VTLP 131
               | ||:|            |   |:|  :             : |:  ::||    ||||
Db        188 FTIDGIKPWGGSLPPNIEGTVYMYAGYYYPM---------KVVYSNAVSWGTLPISVTLP 238

Qy        132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV-DPSAYLYASRVM 190
                  ||  | | |            :|:|       |:     || ||| |   |
Db        239 -----DGTTVSDDFEG--------YVYSFDD-------DLSQSNCTVPDPSNYA-VSTTT 277

Qy        191 PSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESF 250
               :   |  | :   |    :|| | :   : |         |    ||   :| | |
Db        278 TTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTTASTIITTTEPWNSTFTSTST 337

Qy        251 SYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTL 310
               |    :||::       :    :||     |  : : :|   : |
Db        338 ELTTVTGTNGVRTDETII------VIRTPTTATTAITTTEPWNSTFTSTSTEL------- 384

Qy        311 RWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVTTL-PFNPSVDKTKT----------- 355
                    :  |:||:      :   || | :|||:||  |:| :   | |
Db        385 ------TTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGL 438

Qy        356 ------IEILQPIPTTTITT-------SYVGVTTSYLTKTA----PIGETATVIVDVPYH 398
                   | |  |   || |          ::   :|   | |      |  ||  ::: |
Db        439 PTDETIIVIRTPTTATTAMTTTQPWDDTFTSTSTEMTTVTGTNGLPTDET-IIVIRTPTT 497

Qy        399 TTT--TVTSEWTGTIT------TTTTRTN--PTDSIDTVVVQVPLPNPT-VSTTEYWSQS 447
               ||  | |  |  | |          || | ||  |||    :|:: |    | ::||: |: :
Db        498 ATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDE-TIIVIRTPTTATTAMTTTQPWNDT 556

Qy        448 FATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTVTAPP 494
               | :|    ||||   |  ||   ::|| |   |  :|||| |: :| :|    ||||
Db        557 FTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTEPWNSTFTSTSTELTTVTGTN 616

Qy        495 G-GTDS--VIIREPPNPT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPP 542
               |  ||   ::|| |    |  :|||: |: :| :|    ||||   |  ||   ::|| |
Db        617 GLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPT 676

Qy        543 NPT--VTTTEYWSQSYATT----TTVTAPPG-GTDS--VIIREPPNHT--VTTTEYWSQS 591
                  :|||: |: :: :|    ||||   |  ||   ::|| |   |  :|||: |: :
Db        677 TATTAMTTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDT 736

Qy        592 YATT----TTVTAPPG-GTD--TVIIREPPNHTV--TTTEYWSQSFATT----TTVT--- 635
               : :|    ||||   |  ||   ::|| | :  :  |||| |: :| :|    ||||
Db        737 FTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTVTGTN 796

Qy        636 GPPSGTDTVIIREPPNP--TVTTTEYWSQSYATT----TTITAP---PGETDTVLIREPP 686
               | |:    ::|| | :     |||| |: :: :|    ||||     | :   ::|| |
Db        797 GQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMTTITGTNGVPTDETVIVIRTPT 856

Qy        687 NHTV--TTTEYWSQSYATT----TTVTAPPGE-TD--TVLIREPPNHTV--TTTEYWSQS 735
               :  :  |||| |: :: :|    ||:|   |: ||   ::|| | :  :  |||| |: :
Db        857 SEGLISTTTEPWTGTFTSTSTEMTTITGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGT 916

Qy        736 YATT----TTVTAPPG-GTD--TVIIREPPNPTV--TTTEYWSQSFATTTTVTAPPGGT- 785
               : :|    ||||   |  ||   ::|| | :  :  |||| |: :| :|:|      ||
Db        917 FTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEVTTITGTN 976

Qy        786 -----DTVIIYESMSSSKI-----------STSSNDITSIIPSFSRP------------ 816
                    :|||:  : :|  :             :::| ::|::  :  :|
Db        977 GQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPT 1036

Qy        817 --HYVNSTTSD-LSTFESSSMNTPTSISSDGMLLSSTTLVTESETT----TELICSDGKE 869
                  | :||     || |:|    |   ::|:    | :| :: ||    :    | |:
Db       1037 SEGLVTTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSSSGQI 1096

Qy        870 CSRLSSSSGIVTNP--DSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVT 927
                | ::||  |:| |   || :|:::|:| ::|  |   :|:  ||::   : : :  |:
Db       1097 TSSITSSRPIIT-PFYPSNGTSVISSSVISSSVTSSLFTSSPVISSSVISSSTTTSTSIF 1155

Qy        928 TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEI 987
               :|:| :::   |   |:| :| :: ||  ||| |    :| |:  ::| |: | |    :
Db       1156 SESSKSSV--IPTSSSTSGSSESETSSAGSVSSS----SFISSESSKSPTYSSSS----L 1205

Qy        988 EIVTTSST-----KVLPPVVSSNTDLTSEPTNTREQPTTLSTTS-------NSITEDITT 1035
               :||:::|       |||  :         | | || | :: ||        ||:  | :
Db       1206 PLVTSATTSQETASSLPPATT---------TKTSEQTTLVTVTSCESHVCTESISPAIVS 1256

Qy       1036 SQPTGDNGDNTSSTN--PVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTS 1093
               :    :|  | |   |: |  |:     : |     : ::   ::     :  | |
Db       1257 TATVTVSGVTTEYTTWCPISTTETTKQTKGTTEQTTETTKQTTVVTISSCESDVCSKTAS 1316

Qy       1094 TEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTT-------SKTNSELVATT- 1145
                | :|:|:     :   ::   ::|  |::|:| | :: |       |:| |  : :|
Db       1317 PAIVSTSTATINGVTTEYTTWCPIST--TESRQQTTLVTVTSCESGVCSETASPAIVSTA 1374

Qy       1146 ----------------QATNENGGKSPSTDLTSSLTTGT-SASTSANSELVTSGSVTGGA 1188
                              |  ||    |    |   || | :| |: |:    : : || |
Db       1375 TATVNDVVTVYPTWRPQTANEESVSSKMNSATGETTTNTLAAETTTNTVAAETITNTGAA 1434

Qy       1189 -------------------VASASNDQSHSTSV-----TNSNSIVSNTPQTTLSQQVTSS 1224
                                  |||::   ||:||     | :   :::: :|:|||  |:
Db       1435 ETKTVVTSSLSRSNHAETQTASATDVIGHSSSVVSVSETGNTKSLTSSGLSTMSQQPRST 1494

Qy       1225 SP------STNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
                    || :   ||| || : :   : |   |  | | |
Db       1495 PASSMVGYSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1536
```

RESULT 2
US-09-134-001C-4463
; Sequence 4463, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4463
; LENGTH: 2137
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 10.0%; Score 651; DB 4; Length 2137;
Best Local Similarity 24.2%; Pred. No. 3.7e-29;
Matches 314; Conservative 252; Mismatches 541; Indels 192; Gaps 37;

```
Qy         19 TITGV-----FD-SFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMP 72
                |:||:     || | ||:             | ||       ::: | |:
Db        745 TVTGLPQGLKFDASTNSIV-----------GTPTQIGTNTITIESTDASGNK------- 785

Qy         73 CVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPI 132
```

```
Db        786 -----TTTKINYEVTRNSASDSTSTSIVNSVSTSISNST-----SLSDSVKASQSL---- 831

Qy        133 AFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEK--STVDPSAYLYASRVM 190
Db        832 ---------STSKSLSESLSASTSNSTSIQASESASTSKQLSESASTSTSDSASESARKS 882

Qy        191 PSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESF 250
Db        883 ESTSKSTSL-----SESTSTSVSDSASVSTSESASTSTSVSGSTSTSISD---------- 927

Qy        251 SYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFT- 309
Db        928 -STSTSTSDSASIKASESASTSKLLSESVSTSTSDSASTSTSVSDSNSASTSLSKSTSTS 986

Qy        310 ------------LRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTI 356
Db        987 VSDSTSTSTSDSASTSTSESESDSASTSL-SESTSTSVSDSTSTSTSDSASMSASESESN 1045

Qy        357 EILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTT 416
Db       1046 SKSTSLSEST-STSLSGSTSASTSDSA-----STSTSESESDSTSTSLSESTSTSLSGST 1099

Qy        417 RTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTT 476
Db       1100 SASTSDSAST------------STSESDSTSESTSLSESLSTSVSDSTSASTSESASTST 1147

Qy        477 TEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSV 536
Db       1148 SESESNSASTSLSGSLSTSISDSTSTSTSDSASTSTSESESDSTSTSLSESTSTSLSDST 1207

Qy        537 IIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTT 596
Db       1208 STSTSESASTSTSESDSTSESTSLSESTSTSVSDSTSASTSDSASTSTSVSDSESASTSI 1267

Qy        597 TVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTT 656
Db       1268 SESLSTSVSDSTSTSTSDSASTSTSESDSTSESTSLSESISTSVSDSTSASTSDSASTST 1327

Qy        657 TEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTT-------VTAP 709
Db       1328 SE--SESDSASTSLS---GSTSTSL----SDSTSTST---SDSASTSTSESDSERASTSL 1375

Qy        710 PGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWS 769
Db       1376 SGSTSTSL----SDSTSTST---SDSASTSTSVSDSNSASTSL------SGSLSTSVSDS 1422

Qy        770 QSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTF 829
Db       1423 TSTSTSDSASASTSESDSERASTSLSGS-TSTSISDSTSTSTS-------DSASTSTSVS 1474

Qy        830 ESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESS 889
Db       1475 ESNSTSTSISESLSTSVSDSTSTSTSDSASTSTSVSDSDSAS--TSSSESVSTSDSESTS 1532

Qy        890 IVTS-TVPTASTMSDSLSSTDGISATSSDNVSKS-GVSVTTETSVTTIQTTPN--PLSSS 945
Db       1533 TSTSDSASTSTSVSESNSTSTSLSGSTSTSVSDSTSTSTSDSASASTSESDSDSASTSSS 1592

Qy        946 VTSLTQLSSIPSVSESESKVTFTSNGD-NQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSS 1004
Db       1593 ESVSTSVSDSTSASTSESASTSTSVSDSNSASTSLSESTSTSLSDSTSMSTSDSASTSTS 1652

Qy       1005 NTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASAS 1064
Db       1653 ESDSDSASTSLSD------STSTSVSESTSTSTSTSVSASNSTSTSLSDSRSTSLSDSTS 1706

Qy       1065 EEDNKSGS-------HESASTSLKPSMGE--NSGLTTSTEIEATT----------TSPT 1104
Db       1707 TSTSESGSTSTSESDSDSASTSLSESTSTSISDSTSTSTSDSASTSMSVSDSNRASTSLS 1766

Qy       1105 EAPSPAVSSGTDVTTE---PTDTREQPTTLSTTSKTNSELVATTQ-----------ATN 1149
```

```
Db       1767 DSTSTSVSDSTSASTSESASTSTRESESTSASTSLSESTSTSVSDSTSTSTSDSASTSTS 1826

Qy       1150 ENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIV 1209
Db       1827 ESDSNSESTSLSESTSTSVSDSTSAS----TSASASTSTSVSDSNSASTSLSGSTSTSVS 1882

Qy       1210 SNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSI-IQHST 1247
Db       1883 DSTSTSTSASASTSTSESDSDSASTSLSGSTSTSISDST 1921
```

RESULT 3
US-08-362-525-22
; Sequence 22, Application US/08362525
; Patent No. 6027910
; GENERAL INFORMATION:
; APPLICANT: KLIS, FRANCISCUS M.
; APPLICANT: SCHREUDER, MAARTEN P.
; APPLICANT: TOSCHKA, HOLSER Y.
; APPLICANT: VERRIPS, CORNELIS T.
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,525
; FILING DATE: 04-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92202080.5
; FILING DATE: 08-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92203899.7
; FILING DATE: 14-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP93/01763
; FILING DATE: 07-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 213289/T7020(V)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 894 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-362-525-22

Query Match 7.9%; Score 516; DB 3; Length 894;
Best Local Similarity 26.3%; Pred. No. 6.3e-22;
Matches 251; Conservative 126; Mismatches 310; Indels 266; Gaps 43;

```
Qy        437 TVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTT------------------ 477
Db         53 TYSNAAYMAYGYASKTKLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGA 112
```

```
Qy       478 -------EYWSQS-FATTTTVTAPPGGTDSVIIREPPNPTVTTTEY--------WSQSFA 521
                    |||   |   ||              | | |:  | |         ::  ||
Db       113 CSNSQGIAYWSTDLFGFYTT--------------PTNVTLEMTGYFLPPQTGSYTFKFA 157

Qy       522 TT-TTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNH 580
             |   :     ||  :         | :|:| :        |     | ||:       |||
Db       158 TVDDSAILSVGGATAFNCCAQQQPPITSTNF-------TIDGIKPWGGS------LPPNI 204

Qy       581 TVTTTEYWSQSY-------------ATTTTVTAPPGGT-----------------DTVI 609
               |   |               :|| | | |
Db       205 EGTVYMYAGYYYPMKVVYSNAVSWGTLPISVTLPDGTTVSDDFEGYVYSFDDDLSQSNCT 264

Qy       610 IREPPNH----TVTTTEYWSQSFATT----TTVTGP---PSGTDTVIIREPPNPTV--TT 656
             : :| |:    | |||| |: :| :|    |||||    |:    ::|| | :  :  ||
Db       265 VPDPSNYAVSTTTTTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTT 324

Qy       657 TEYWSQSYATT----TTITAPPGE-TD--TVLIREPPNHTV--TTTEYWSQSYATT---- 703
             || |: :: :|    ||||   |: ||   ::|| | :  :  |||| |: :: :|
Db       325 TEPWTGTFTSTSTEVTTITGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEM 384

Qy       704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNP--T 761
             ||||  |        :| : ||                          ::|| | :
Db       385 TTVTGTNG--------QPTDETV------------------------IVIRTPTSEGLV 411

Qy       762 VTTTEYWSQSFATTTTVTAPPGGT------DTVIIYE----SMSSSKISTSSNDITSIIP 811
             |||| |: :| :|:|   :   ||      :|||: :    ::||| |:|| |||  | |
Db       412 TTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSSSGQITSSIT 471

Qy       812 SFSR----PHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDG 867
             | ||    | | :: || :|:   ||  | :  :|  ::  |:::: | ||:  | |:
Db       472 S-SRPIITPFYPSNGTSVISSSVISSSVTSSLFTSSPVI--SSSVISSSTTTSTSIFSES 528

Qy       868 KECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVT 927
             : | : :||    : :|  ||            : |:||:  ||: || : : |  |:
Db       529 SKSSVIPTSSSTSGSSESETSS-----------AGSVSSSSFISSESSKSPTYSSSSLP 576

Qy       928 TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEI 987
               || || | | : |  : |: |        ||  : ||                 || |
Db       577 LVTSATTSQETASSLPPATTTKT--------SEQTTLVTV--------------TSCES 613

Qy       988 EIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTS 1047
             : | |   : | :||: |   |  |        :|||        || |  |   : |:
Db       614 HVCTES---ISPAIVSTATVTVSGVTTEYTTWCPISTTE-------TTKQTKGTT-EQTT 662

Qy      1048 STNPVPTVATSTLASASEED--NKSGSHESASTSLKPSMGENSGLTT-----STEIEATT 1100
               |   || |   :| | |  :|: |   |  |||     |  : ||     :||     |
Db       663 ETTKQTTVVT---ISSCESDVCSKTASPAIVSTSTATINGVTTEYTTWCPISTTESRQQT 719

Qy      1101 TSPT----------EAPSPA-VSSGTD-----VTTEPTDTREQPTTLSTTSKTNSELVAT 1144
               |  |          |  ||| ||: |      ||  ||   :    | :|| ||    |
Db       720 TLVTVTSCESGVCSETASPAIVSTATATVNDVVTVYPTWRPQTANEESVSSKMNSATGET 779

Qy      1145 TQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGG-----AVASASNDQSHS 1199
               |  ||    :: :  : :   | | |   | :: | : |::         |||::   ||
Db       780 T--TNTLAAETTTNTVAAETITNTGA---AETKTVVTSSLSRSNHAETQTASATDVIGHS 834

Qy      1200 TSV-----TNSNSIVSNTPQTTLSQQVTSSSP------STNTFIASTYDGSGS 1241
               :||     | :   ::::  :|:|||  |:      || :   ||| || :
Db       835 SSVVSVSETGNTKSLTSSGLSTMSQQPRSTPASSMVGYSTASLEISTYAGSAT 887


RESULT 4
US-08-971-692-15
; Sequence 15, Application US/08971692
; Patent No. 6114147
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Immobilized proteins with specific binding
;    TITLE OF INVENTION:  capacities and their use in processes and products.
;    NUMBER OF SEQUENCES:  40
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/971,692
;      FILING DATE:
;      CLASSIFICATION:  435
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  894 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-971-692-15

  Query Match             7.9%;  Score 516;  DB 3;  Length 894;
  Best Local Similarity   26.3%;  Pred. No. 6.3e-22;
  Matches  251;  Conservative  126;  Mismatches  310;  Indels  266;  Gaps   43;

Qy       437 TVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTT------------------- 477
             | |   | :  :|: | : :  | ||  |    |  : : |
Db        53 TYSNAAYMAYGYASKTKLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGA 112

Qy       478 -------EYWSQS-FATTTTVTAPPGGTDSVIIREPPNPTVTTTEY--------WSQSFA 521
                    |||   |   ||              | | |:  | |         ::  ||
Db       113 CSNSQGIAYWSTDLFGFYTT--------------PTNVTLEMTGYFLPPQTGSYTFKFA 157

Qy       522 TT-TTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNH 580
             |   :     ||  :         | :|:| :        |     | ||:       |||
Db       158 TVDDSAILSVGGATAFNCCAQQQPPITSTNF-------TIDGIKPWGGS------LPPNI 204

Qy       581 TVTTTEYWSQSY-------------ATTTTVTAPPGGT-----------------DTVI 609
               |   |               :|| | | |
Db       205 EGTVYMYAGYYYPMKVVYSNAVSWGTLPISVTLPDGTTVSDDFEGYVYSFDDDLSQSNCT 264

Qy       610 IREPPNH----TVTTTEYWSQSFATT----TTVTGP---PSGTDTVIIREPPNPTV--TT 656
             : :| |:    | |||| |: :| :|    |||||    |:    ::|| | :  :  ||
Db       265 VPDPSNYAVSTTTTTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTT 324

Qy       657 TEYWSQSYATT----TTITAPPGE-TD--TVLIREPPNHTV--TTTEYWSQSYATT---- 703
             || |: :: :|    ||||   |: ||   ::|| | :  :  |||| |: :: :|
Db       325 TEPWTGTFTSTSTEVTTITGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEM 384

Qy       704 TTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNP--T 761
             ||||  |        :| : ||                          ::|| | :
Db       385 TTVTGTNG--------QPTDETV------------------------IVIRTPTSEGLV 411

Qy       762 VTTTEYWSQSFATTTTVTAPPGGT------DTVIIYE----SMSSSKISTSSNDITSIIP 811
             |||| |: :| :|:|   :   ||      :|||: :    ::||| |:|| |||  | |
Db       412 TTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSSSGQITSSIT 471

Qy       812 SFSR----PHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDG 867
             | ||    | | :: || :|:   ||  | :  :|  ::  |:::: | ||:  | |:
Db       472 S-SRPIITPFYPSNGTSVISSSVISSSVTSSLFTSSPVI--SSSVISSSTTTSTSIFSES 528

Qy       868 KECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVT 927
             : | : :||    : :|  ||            : |:||:  ||: || : : |  |:
Db       529 SKSSVIPTSSSTSGSSESETSS-----------AGSVSSSSFISSESSKSPTYSSSSLP 576

Qy       928 TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEI 987
               || || | | : |  : |: |        ||  : ||                 || |
Db       577 LVTSATTSQETASSLPPATTTKT--------SEQTTLVTV--------------TSCES 613

Qy       988 EIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTS 1047
             : | |   : | :||: |   |  |        :|||        || |  |   : |:
Db       614 HVCTES---ISPAIVSTATVTVSGVTTEYTTWCPISTTE-------TTKQTKGTT-EQTT 662

Qy      1048 STNPVPTVATSTLASASEED--NKSGSHESASTSLKPSMGENSGLTT-----STEIEATT 1100
               |   || |   :| | |  :|: |   |  |||     |  : ||     :||     |
Db       663 ETTKQTTVVT---ISSCESDVCSKTASPAIVSTSTATINGVTTEYTTWCPISTTESRQQT 719

Qy      1101 TSPT----------EAPSPA-VSSGTD-----VTTEPTDTREQPTTLSTTSKTNSELVAT 1144
               |  |          |  ||| ||: |      ||  ||   :    | :|| ||    |
Db       720 TLVTVTSCESGVCSETASPAIVSTATATVNDVVTVYPTWRPQTANEESVSSKMNSATGET 779
```

```
Qy       1145 TQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGG-----AVASASNDQSHS 1199
Db        780 T--TNTLAAETTTNTVAAETITNTGA---AETKTVVTSSLSRSNHAETQTASATDVIGHS 834

Qy       1200 TSV-----TNSNSIVSNTPQTTLSQQVTSSSP------STNTFIASTYDGSGS 1241
Db        835 SSVVSVSETGNTKSLTSSGLSTMSQQPRSTPASSMVGYSTASLEISTYAGSAT 887


RESULT 5
US-08-928-361B-6
; Sequence 6, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-1677
;      TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1721 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-928-361B-6

  Query Match             7.5%;  Score 488;  DB 3;  Length 1721;
  Best Local Similarity   22.5%;  Pred. No. 6.2e-20;
  Matches  303;  Conservative  128;  Mismatches  469;  Indels  448;  Gaps   58;

Qy        214 MGFSSSNGDVAIDCSNIHIGITKG-LNDWNYPVSSESFSYTKTC---------TSNGIQI 263
Db         19 LNFTSTTGFTTDTSMNWPVSITSGELKD---PNKQATISGSRSCGWKQGYSIDSSTGFRV 75

Qy        264 ---------KYQNVPAGYRPFIDAYIS----ATDVNQYTLAYTNDYTCAGSRLQSKPFTL 310
Db         76 DSITGLPTDPYSNCP--FNPVTGNLVSRSTGKTIPNTYAGVYRSNET-----KTTEPSAN 128

Qy        311 RWTG-YKNSD-------AGSNGIVI---------------------------------- 327
Db        129 TYAGVYRSNETKTTEPSANTNFLLVDPKINAPCNSENSFEQGQIFDMGSKVYIPYTKCVG 188

Qy        328 ----------VATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTS 377

Db        189 VKHTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 248

Qy        378 YLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTTRTNPTDSIDTVVVQVPLPNPT 437
Db        249 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPT 308

Qy        438 VSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGT 497
Db        309 TTTTT-TTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTT--------TTTTTTTTTTT 359

Qy        498 DSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYA 557
Db        360 ----------TTTTTTTTTTTTTTTTTTTKPTTTTTTTTTTTTTKKPTTTTT-------A 401

Qy        558 TTTTVTAPPGGTDSVIIRE-----------------------------------PPNH 580
Db        402 TTTTTTSE---TESVIKPDEWCWLEKNGECEAKGATYVGVIGKDGRIENGMAFTMIPNDD 458

Qy        581 TVTTTEYWSQSYATTTTV-----------------TAPP------------------- 602
Db        459 THVRFRFKVKDVGNTISVRCRKGAGKLEFPDRSLDFTIPPVAGHNSCSIIVGVSGDGKIH 518

Qy        603 ----GGTDTVIIREP--PNH-------TVTTTEY--------WSQSFAT------TTTVT 635
Db        519 VSPYGSKDVSLISAPIQPSELFNEVYCDTCTAKYGAIHSGYQTSADFVTTTTAKPTTTTT 578

Qy        636 GPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEY 695
Db        579 GAPGQPTTTTTGSPSKPTTTTT-----TKATTTTTTLNP--------------IITTT-- 617

Qy        696 WSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIR 755
Db        618 --TQKPTTTTTTKVPG--------KPPIATTTTT---LKPIVTTTTTKATTTTTTV--- 661

Qy        756 EPPNPTVTTTEYWSQSFATTTTVTAPPGG--------------TDTVIIYESMSSSKIST 801
Db        662 ----PTTTTTT--KRDEMTTTTTPLPDIGDIEITPIPIEKMLDKYTRMIYDYNSGLLLDS 715

Qy        802 SSN-----------DITSIIP----------------------SFSRPH-YVNSTTSDLS 827
Db        716 NDEPIPGSQAGQIADTSNLFPVQTHKSTGLPIDPMVGLPFDPKSGNLVHPYTNQTMSGLS 775

Qy        828 TFESSSMNTPTSISSD-----------GMLLSSTTLVTESETTTELICSDGKECSRLSSS 876
Db        776 VSYLAAKN--LTVDTDETYGLPIDTLTGYPLDPVSLIPFNPETGELFDPISDEIMN-GTI 832

Qy        877 SGIVTNPDSNES--SIVTSTVPTASTM-----SDSLSSTDGI----------SATSSDNV 919
Db        833 AGIVSGISASESLLSQKSALIDPATNMVVGEFGGLLNPATGVMIPGFLGPSEQTQFSPEI 892

Qy        920 SKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSES----ESKVTFTSN--GDN 973
Db        893 EDGGI-------------IPPEVAAANADKFKLSIPPSVPESIPEKDQKIDSISELMYDI 939

Qy        974 QSGTHDSQ----------------STSTEIEIVT------TSSTKVLPPVVSSNTDLT 1009
Db        940 ESGRLIGQVSKRPIPGSIAGDLNPIMKTPTQTDSVTGKPIDPTTGLPFNPP-----TGHL 994

Qy       1010 SEPTNTREQPTTLS-----TTSNSITEDITTSQPTGD-NGDNTSSTNPVPTVATS----T 1059
Db        995 INPTNNNTMDSSFAGAYKYAVSNGIKTDNVYGLPVGEITGLPKDPGSDIPFNSTTGELVD 1054

Qy       1060 LASASEEDNKSGSHESASTSLKPSMGENSGL---TTSTEIEATTTSPTEAPSPAVSSGTD 1116
Db        1055 PSTGKPINNSTAGIVSGKPGLPPIEDENGNLFDPSTNLPIDGNNQLVNPETNSTVSGSTS 1114

Qy       1117 VTTEP--------------TDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTS 1162
Db        1115 GTTKPKPGIPVNGGGVVPDEEAKDQ-----ADKGKDGLIVPPTNSIN----KDPVTNTQY 1165

Qy       1163 SLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVT 1222
```

```
Db       1166 SNTTG----NIINPE---TGKVIPGSLPGSLNYPSF-----------NTPQQT--DEIT 1204

Qy       1223 SSSPSTNTFIASTYDGS-GSIIQHSTWL 1249
                 | | :   || | | ||  :| |
Db       1205 GKPVDTVTGL--PYDPSTGEIIDPATKL 1230


RESULT 6
US-08-700-651-5
; Sequence 5, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
;  APPLICANT: PETERSEN, CAROLYN
;  APPLICANT: LEECH, JAMES
;  APPLICANT: NELSON, RICHARD, C.
;  APPLICANT: GUT, JIRI
;  TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
;  TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
;  TITLE OF INVENTION: INFECTIONS
;  FILE REFERENCE: 480.19-4(HV)
;  CURRENT APPLICATION NUMBER: US/08/700,651B
;  CURRENT FILING DATE: 1997-08-14
;  EARLIER APPLICATION NUMBER: 08/415,751
;  EARLIER FILING DATE: 1995-04-03
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1721
;   TYPE: PRT
;   ORGANISM: Cryptosporidium parvum
US-08-700-651-5

  Query Match             7.5%;  Score 487;  DB 3;  Length 1721;
  Best Local Similarity   22.3%;  Pred. No. 7.1e-20;
  Matches  301;  Conservative  129;  Mismatches  471;  Indels  446;  Gaps   57;

Qy        214 MGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTC---------TSNGIQI- 263
              : |:|: |       |  : || |  :   |    : | :::|         :| | ::
Db         19 LNFTSTTGFTTDTSMNWPVSITSG--ERKDPNKQATISGSRSCGWKQGYSIDSSTGFRVD 76

Qy        264 --------KYQNVPAGYRPFIDAYIS----ATDVNQYTLAYTNDYTCAGSRLQSKPFTLR 311
                     | | |  : |     :|     |  | |   | :: |       ::|
Db         77 SITGLPTDPYSNCP--FNPVTGNLVSRSTGKTIPNTYAGVYRSNET-----KTTEPSANT 129

Qy        312 WTG-YKNSD-------AGSNGIVI---------------------------------- 327
              : | |::::       | :| :::
Db        130 YAGVYRSNETKTTEPSANTNFLLVDPKINAPCNSENSFEQGQIFDMGSKVYIPYTKCVGV 189

Qy        328 ---------VATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSY 378
                      || | | :||  ||     :   | |         ||| ||:    ||:
Db        190 KHTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 249

Qy        379 LTKTAPIGETATVIVDVPYHTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTV 438
               | |     | |        ||||  |:  | | |||||  |  | :  |          ||
Db        250 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTT 309

Qy        439 STTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTD 498
              :||   : :  |||| |     | |   ::|   | |||        |||| |     |
Db        310 TTTT-TTTTTTTTTTTTTTTTTTTTTTKKPTTTTTTTT--------TTTTTTTTTTT- 359

Qy        499 SVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYAT 558
                    | |||    : :  |||| | |   | :         || |||       ||
Db        360 ----------TTTTTTTTTTTTTTTTTTTKPTTTTTTTTTTTTKKPTTTT-------AT 402

Qy        559 TTTVTAPPGGTDSVIIRE-----------------------------------PPNHT 581
              |||  |:    |:|||  :                                  | : |
Db        403 TTTTTSE---TESVIKPDEWCWLEKNGECEAKGATYVGVIGKDGRIENGMAFTMIPNDDT 459

Qy        582 VTTTEYWSQSYATTTTV-----------------TAPP---------------------- 602
              :  :    | :|                  | ||
Db        460 HVRFRFKVKDVGNTISVRCRKGAGKLEFPDRSLDFTIPPVAGHNSCSIIVGVSGDGKIHV 519

Qy        603 ---GGTDTVIIREP--PNH-------TVTTTEY--------WSQSFAT------TTTVTG 636
                    | |  :|  |  |:           | :|         |  | |      ||| ||
Db        520 SPYGSKDVSLISAPIQPSELFNEVYCDTCTAKYGAIHSGYQTSADFVTTTTAKPTTTTTG 579

Qy        637 PPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYW 696
               |    |     |  || |||     : |||||  |  |               :|||
Db        580 APGQPTTTTTGSPSKPTTTTT-----TKATTTTTTLNP--------------IITTT--- 617

Qy        697 SQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIRE 756
                  |||| |  ||        :||  | |||    :   ||||  |    | ||
Db        618 -TQKPTTTTTTKVPG--------KPPIATTTTT---LKPIVTTTTTKATTTTTTV---- 661

Qy        757 PPNPTVTTTEYWSQSFATTTTVTAPPGG-------------TDTVIIYESMSSSKISTS 802
                 || |||    :  ||||   |  |                | :||:  |   : ::
Db        662 ---PTTTTTT--KRDEMTTTTTPLPDIGDIEITPIPIEKMLDKYTRMIYDYNSGLLLDSN 716

Qy        803 SN-----------DITSIIP---------------------SFSRPH-YVNSTTSDLST 828
                          | ::: |                     | :  | | | | | ||
Db        717 DEPIPGSQAGQIADTSNLFPVQTHKSTGLPIDPMVGLPFDPKSGNLVHPYTNQTMSGLSV 776

Qy        829 FESSSMNTPTSISSD----------GMLLSSTTLVTESETTTELICSDGKECSRLSSSS 877
                ::  |   :: :|          |  |   :|:  :  | ||      |    : :
Db        777 SYLAAKN--LTVDTDETYGLPIDTLTGYPLDPVSLIPFNPETGELFDPISDEIMN-GTIA 833

Qy        878 GIVTNPDSNES--SIVTSTVPTASTM-----SDSLSSTDGI----------SATSSDNVS 920
                |||:   ::||  |  :: :  |: |        |:   |:           |  :
Db        834 GIVSGISASESLLSQKSALIDPATNMVVGEFGGLLNPATGVMIPGFLGPSEQTQFSPEIE 893

Qy        921 KSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSES----ESKVTFTSN--GDNQ 974
                |:              |  :::     :||  ||| ||    : |:   |    | :
Db        894 DGGI------------IPPEVAAANADKFKLSIPPSVPESIPEKDQKIDSISELMYDIE 940

Qy        975 SGTHDSQ-----------------STSTEIEIVT------TSSTKVLPPVVSSNTDLTS 1010
                ||    |                 | |: : ||       |:     ||     |
Db        941 SGRLIGQVSKRPIPGSIAGDLNPIMKTPTQTDSVTGKPIDPTTGLPFNPP-----TGHLI 995

Qy       1011 EPTNTREQPTTLS-----TTSNSITEDITTSQPTGD-NGDNTSSTNPVPTVATS----TL 1060
                |||     :: :       || |  |     | |: |     : .:|  :|:
Db        996 NPTNNNTMDSSFAGAYKYAVSNGIKTDNVYGLPVGEITGLPKDPGSDIPFNSTTGELVDP 1055

Qy       1061 ASASEEDNKSGSHESASTSLKPSMGENSGL---TTSTEIEATTTSPTEAPSPAVSSGTDV 1117
                ::    :| :    |  | |   || |   :|:  |:          :  || |
Db       1056 STGKPINNSTAGIVSGKPGLPPIEDENGNLFDPSTNLPIDGNNQLVNPETNSTVSGSTSG 1115

Qy       1118 TTEP-------------TDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSS 1163
                ||:|              : ::|          :  :|  | : |    | | |:   |
Db       1116 TTKPKPGIPVNGGGVVPDEEAKDQ-----ADKGKDGLIVPPTNSIN----KDPVTNTQYS 1166

Qy       1164 LTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTS 1223
                |||       | |  :| |  |::  : | |              ||||  |   ::|
Db       1167 NTTG----NIINPE---TGKVIPGSLPGSLNYPSF-----------NTPQQT--DEITG 1205

Qy       1224 SSPSTNTFIASTYDGS-GSIIQHSTWL 1249
                  | | :   || | | ||  :| |
Db       1206 KPVDTVTGL--PYDPSTGEIIDPATKL 1230


RESULT 7
US-08-928-361B-5
; Sequence 5, Application US/08928361B
; Patent No. 6071518
;  GENERAL INFORMATION:
;    APPLICANT:  Petersen, Carolyn
;    TITLE OF INVENTION:  PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
;    TITLE OF INVENTION:  THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
;    TITLE OF INVENTION:  FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
;    TITLE OF INVENTION:  SPECIES INFECTIONS
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PETERS, VERNY, JONES & BIKSA
;      STREET:  385 Sherman Avenue, Suite 6
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
```

```
;      ZIP:  94306-1840
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/928,361B
;      FILING DATE:  12-SEP-1997
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/026,062
;      FILING DATE:  13-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verny, Hana
;      REGISTRATION NUMBER:  30,518
;      REFERENCE/DOCKET NUMBER:  480.76-1(HV)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-324-1677
;      TELEFAX:  650-324-1678
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1837 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-928-361B-5

  Query Match             7.5%;  Score 487;  DB 3;  Length 1837;
  Best Local Similarity   22.8%;  Pred. No. 7.8e-20;
  Matches  316;  Conservative  146;  Mismatches  487;  Indels  440;  Gaps   61;

Qy        136 VGGTGSSTDLEDSKCFTAGTNTVTFN-DGDKDISIDVEFEKST---VDP-SAYLY--ASR 188
              :  ||         |   | | :  ::   : |:|  |  | :|   ||| |  |:  ::
Db         69 IDSTGKKFSPYTGKHADASTTSSAYSAPFELDVS-GVPIEPNTRRMVDPVSLMLFDNSTG 127

Qy        189 VM--PSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVS 246
              ||  |: | :    :|         : |:|: |       |  : || |
Db        128 VMYDPNTNSILEGSIAGIRSESCIVSELNFTSTTGFTTDTSMNWPVSITSG--------- 178

Qy        247 SESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSK 306
                          ::|  |  |                          | :|||
Db        179 ---------------ELKDPNKQA-------------------------TISGSRS--- 194

Qy        307 PFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTT-----LPFNPSVDKTKTIEILQP 361
                     |:|       |  | ::|    || | : |      ||||       |  ::
Db        195 ------CGWK------QGYSIDSSTGFRVDSITGLPTDPYPNCPFNP-----VTGNLVSR 237

Qy        362 IPTTTITTSYVGVTTSYLTKTA-PIGETATVIVD------------------------- 394
                  ||  :| ||  |  |||  |   |  ::||
Db        238 STGKTIPNTYAGVYRSNETKTTEPSANTNFLLVDPKINAPCNSENSFEQVQIFDMGSKVY 297

Qy        395 VPY-------HTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQS 447
              :||       ||||| |:  | | |||||  |  | :  |           | :||   : :
Db        298 IPYTKCVGVKHTTTTTTTT-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 356

Qy        448 FATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPN 507
                ||||  |      | |   ::|   | |||   : :  ||||
Db        357 TTTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTT--------------T 401

Qy        508 PTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPG 567
              | |||   : :  ||||  |  |  | :          | |||   : :  |||| |  |
Db        402 TTTTTTTTTTTTTTTTTTTKKPTTTTTT----TTTTTTTTTTTTTTTTTTTTTTKKPT 457

Qy        568 GTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQS 627
              | :          | |||   : :  |||| ||       |   ::|   | |||   ::
Db        458 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTA------TTTTKKPTTTTTTTTTTTKK 511

Qy        628 FATTTTVTGPPSGTDTVIIREP------------------------------------ 649
                ||||  |   : ::|  : :|
Db        512 PTTTTTATTTTTTSETESVIKPDEWCWLEKNGECEAKGATYVGVIGKDGRIENGMAFTMI 571

Qy        650 PN-PTVTTTEYWSQSYATTTTI-----------------TAPP--------------- 674
              ||  |      :  :    | ::                  | ||
Db        572 PNDDTHVRFRFKVKDVGNTISVRCGKGAGKLEFPDRSLDFTIPPVAGHNSCSIIVGVSGG 631

Qy        675 --------GETDTVLIREPPNHTVTTTEYWSQSYATT-----------------TTVT 707
                     |  |  ||  |    :  | ::: |  |                    || |
Db        632 GKIHVSPYGSKDVSLISAP----IQPCELFNEVYCDTCTAKYGAIHSGYQTSADFVTTTT 687

Qy        708 APPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEY 767
              | |  | |    :|     |||   | |  |||| |     | | |:    || :|||
Db        688 AKPTTTTTGAPGQP-----TTTTTGSPSKPTTTTTT--KATTTTTIL----NPIITTT-- 734

Qy        768 WSQSFATTTTVTA---PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTS 824
               :|   ||||      ||  | |  :   ::::    ::   |:          | :||:
Db        735 -TQKPTTTTTTKVPGKPPIATTTTTLKPIVTTTTTKATTTTTTT----------VPTTTT 783

Qy        825 DLSTFESSSMNTP---------TSISSDGMLLSSTTLVTESETTTELICSD----GKECS 871
                  | ::  ||         | |  : ||   | :: :  :   |  :|    | :
Db        784 TTKRDEMTTTTTPLPDIGDIEITPIPIEKMLDKYTRMIYDYNSGLLLDSNDEPIPGSQAG 843

Qy        872 RLS-----------SSSGIVTNP------DSNESSIV-TSTVPTASTMSDSLSSTDGISA 913
              :::            |:|:  :|      |    ::|   |  | | :| |  :   ::
Db        844 QIADTSNLFPVQTHKSTGLPIDPMVGLPFDPKSGNLVHPYTNQTMSGLSVSYLAAKNLTV 903

Qy        914 TSSDNVSKSGVSVTTETS-----VTTIQTTP------NPLSSSVTSLTQLSSIPSVSESE 962
               : :     |: : | |      |: |   |      :|:|  : : |    :  :| ||
Db        904 DTDETY---GLPIDTLTGYPLDPVSLIPFNPETGELFDPISDEIMNGTIAGIVSGISASE 960

Qy        963 SKVTFTSN----------GD-------------NQSGTHDSQSTSTEIEIVTTSSTKVL 998
              | ::  |           |:                  |  :    | |||        ::
Db        961 SLLSQKSAPIDPATNMVVGEFGGLLNPATGVMIPGSLGPSEQTPFSPEIE-----DGGII 1015

Qy        999 PP-VVSSNTD--LTSEPTNTREQPTTLSTTSNSITE---DITTSQPTGD------NGDNT 1046
              || | ::| |    | | :  |          :||:|   || : :  |          |
Db       1016 PPEVAAANADKFKLSIPPSVPESIPEKDQKIDSISELMYDIESGRLIGQVSKRPIPGSIA 1075

Qy       1047 SSTNPVPTVATSTLASASEEDNKSGSHESASTSL--KPSMGENSGLTTSTEIEATTTSPT 1104
               ||:    | |       |: :|     :| |   |  |   |    |  |      :|
Db       1076 GDLNPIMKTPTQT-------DSVTGKPIDPTTGLPFNPPTGH---LINPTNNNTMDSSFA 1125

Qy       1105 EAPSPAVSSG--TD-----VTTEPTDTREQPTTLSTTSKTNSELV--ATTQATN------ 1149
               |   |||:|  ||       | |   : | :    : |  |||  :| :  |
Db       1126 GAYKYAVSNGIKTDNVYGLPVDEITGLPKDPVSDIPFNSTTGELVDPSTGKPINNYTAGI 1185

Qy       1150 ------------ENGGK-SPSTDL------------TSSLTTG-TSASTSANSELVTSGS 1183
                          |||    ||| |            |:|  :| || ||     :  :|
Db       1186 VSGKRGLPPIEDENGNLFDPSTKLPIDGNNQLVNPETNSTVSGSTSGSTKPKPGIPVNG- 1244

Qy       1184 VTGGAVA-SASNDQSHSTS----VTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDG 1238
               || |    : ||:        |  :|||              :  | |||  ::|
Db       1245 --GGVVPDEEAKDQADKGKDGLIVPPTNSI--------------NKDPVTNTQYSNT--- 1285

Qy       1239 SGSIIQHST 1247
              :|:||   |
Db       1286 TGNIINPET 1294


RESULT 8
US-08-325-267A-4
; Sequence 4, Application US/08325267A
; Patent No. 5585271
;  GENERAL INFORMATION:
;    APPLICANT:  WATARI, JUNJI
;    APPLICANT:  TAKATA, YOSHIHIRO
;    APPLICANT:  OGAWA, MASAHIRO
;    APPLICANT:  PENTTILA, MERJA
;    APPLICANT:  ONNELA, MAIJA-LEENA
;    APPLICANT:  KERANEN, SIRKKA
;    TITLE OF INVENTION:  YEAST AGGLUTINATION GENES AND YEAST
;    TITLE OF INVENTION:  CONTAINING THEM
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
```

```
;      STREET:  1755 S. JEFFERSON DAVIS HWY, SUITE 400
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0; Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/325,267A
;      FILING DATE:  18-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP PCT/JP94/00290
;      FILING DATE:  24-FEB-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 38871/1993
;      FILING DATE:  26-FEB-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2589-023-0XPCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 862 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-325-267A-4

  Query Match             7.2%;  Score 468.5;  DB 1;  Length 862;
  Best Local Similarity   25.1%;  Pred. No. 3.2e-19;
  Matches  241;  Conservative  120;  Mismatches  313;  Indels  285;  Gaps   39;

Qy        437 TVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTT------------------ 477
              | |   | :  :|: | : :  | ||  |    |  : : |
Db         53 TYSNAAYMAYGYASKTKLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGA 112

Qy        478 -------EYWSQS-FATTTTVTAPPGGTDSVIIREPPNPTVTTTEY--------WSQSFA 521
                      |||   |   ||                | | |:  | |       ::  ||
Db        113 CSNSQGIAYWSTDLFGFYTT--------------PTNVTLEMTGYFLPPQTGSYTFKFA 157

Qy        522 TT-TTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNH 580
              |   :     ||  :        | :|:| :       |     | ||:       |||
Db        158 TVDDSAILSVGGATAFNCCAQQQPPITSTNF-------TIDGIKPWGGS------LPPNI 204

Qy        581 TVTTTEYWSQSY------------ATTTTVTAPPGGT------------------DTVI 609
                |   |    |                :|| | | |
Db        205 EGTVYMYAGYYYPMKVVYSNAVSWGTLPISVTLPDGTTVSDDFEGYVYSFDDDLSQSNCT 264

Qy        610 IREPPNH----TVTTTEYWSQSFATT----TTVTGP---PSGTDTVIIREPPNPT--VTT 656
              : :| |:    | |||| |: :| :|    ||||     |:    ::|| |   :  :||
Db        265 VPDPSNYAVSTTTTTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTTASTIITT 324

Qy        657 TEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTV 716
              || |: :: :|:|                                    ||||   |
Db        325 TEPWTGTFTSTST------------------------------EMTTVTGTNG----- 347

Qy        717 LIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNP--TVTTTEYWSQSFAT 774
                 :| : ||                          ::|| | :     |||| |: :| :
Db        348 ---QPTDETV------------------------IVIRTPTSEGLVTTTTEPWTGTFTS 379

Qy        775 TTTVTAPPGGT------DTVIIYE----SMSSSKISTSSNDITSIIPSFSR----PHYVN 820
              |:|   :   ||      :|||: :    ::|||  |:||  ||| | | ||    | | :
Db        380 TSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSSSGQITSSITS-SRPIITPFYPS 438

Qy        821 STTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIV 880
```

```
              : || :|:   ||  | :  :|  ::  |:::::| ||:  | |:  : | : :||
Db        439 NGTSVISSSVISSSVTSSLFTSSPVI--SSSVISSSTTTSTSIFSESSKSSVIPTSSSTS 496

Qy        881 TNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPN 940
               : :|  ||            : |:||:  ||: || : : |  |:   || || | | :
Db        497 GSSESETSS-----------AGSVSSSSFISSESSKSPTYSSSSLPLVTSATTSQETAS 544

Qy        941 PLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPP 1000
               | : |: |          ||  : ||                  || |  : | |   : |
Db        545 SLPPATTTKT--------SEQTTLVTV--------------TSCESHVCTES---ISPA 578

Qy       1001 VVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTL 1060
               :||: |   |  |         :|||        || |  |   : |: |    || |
Db        579 IVSTATVTVSGVTTEYTTWCPISTTE-------TTKQTKGTT-EQTTETTKQTTVVT--- 627

Qy       1061 ASASEED--NKSGSHESASTSLKPSMGENSGLTT-----STEIEATTTSPT--------- 1104
               |: | |  :|: |    |||      |  :  ||     :||    ||  |
Db        628 ISSCESDVCSKTASPAIVSTSTATINGVTTEYTTWCPISTTESRQQTTLVTVTSCESGVC 687

Qy       1105 -EAPSPA-VSSGTD-----VTTEPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPS 1157
               |  ||| ||: |       ||  ||   :      | :|| ||    ||  ||    :: :
Db        688 SETASPAIVSTATATVNDVVTVYPTWRPQTANEESVSSKMNSATGETT--TNTLAAETTT 745

Qy       1158 TDLTSSLTTGTSASTSANSELVTSGSVTGG-----AVASASNDQSHSTSV-----TNSNS 1207
                 : :   | | |   | :: | : |::         |||::   ||:||     | :
Db        746 NTVAAETITNTGA---AETKTVVTSSLSRSNHAETQTASATDVIGHSSSVVSVSETGNTK 802

Qy       1208 IVSNTPQTTLSQQVTSSSP-----STNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
               ::::  :|:|||  |:        || :  ||| || : :   : |   |  | | |
Db        803 SLTSSGLSTMSQQPRSTPASSMVGYSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 861
```

```
RESULT 9
US-08-046-585-5
; Sequence 5, Application US/08046585
; Patent No. 5453362
;  GENERAL INFORMATION:
;    APPLICANT:  Lamarco, Kelly
;    APPLICANT:  Wilson, Angus
;    APPLICANT:  Herr, Winship
;    TITLE OF INVENTION:  A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
;    TITLE OF INVENTION:  HOST CELL FACTOR
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/046,585
;      FILING DATE:  12-APR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-57503-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2035 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  peptide
US-08-046-585-5

  Query Match             6.2%;  Score 404.5;  DB 1;  Length 2035;
  Best Local Similarity   21.9%;  Pred. No. 4.9e-15;
  Matches  319;  Conservative  165;  Mismatches  506;  Indels  467;  Gaps   69;

Qy         14 IASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAV-LGWSLDGTSA-NPGDTFTLNM 71
              | |:  ::|:     :|  : ||          ||  :|  | ::  | |  ||   |  :
Db        537 IGSSPQMSGM----AALAAAAAATQKIPPSSAPTVLSVPAGTTIVKTMAVTPG---TTTL 589

Qy         72 PCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLP 131
              |   |  :|   |              |       |     | :: |   |  | | |
Db        590 PATVKVASSPVMV--------------SNPATRMLKTAAAQVGTSVSS---ATNTSTRP 631

Qy        132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMP 191
              |   |  :|: |  : ::  |     ||     | |::          | :    | ::
Db        632 I-ITVHKSGTVTVAQQAQVVTTVVGGVT-----KTITL-------VKSPISVPGGSALIS 678

Qy        192 SLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFS 251
              :| || ::    | :   ||   |  :| | |        |  |||
Db        679 NLGKVMSVV---QTKPVQTSAVTG-QASTGPVT------QIIQTKG-------------- 714

Qy        252 YTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLR 311
                          :|||    |   ::: |    |:  |   : ||    :||  |
Db        715 ---------------PLPAG---TILKLVTSADGKPTTIITTTQASGAG----TKPTIL- 751

Qy        312 WTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDK------TKTIEILQPIP-- 363
                        ||  |: : |   :|| : |:| :  : :       | :  |  ||
Db        752 -----------GISSVSPSTTKPGTTTIIKTIPMSAIITQAGATGVTSSPGIKSPITII 799

Qy        364 TTTIITTSYVGVTTSYLT------------------KTAPIGETATVIVDVPYHTTTTVT 404
              || : ||  |     :|                  | || |:  |::  ||       ||
Db        800 TTKVMTSGTGAPAKIITAVPKIATGHGQQGVTQVVLKGAP-GQPGTILRTVPMGGVRLVT 858

Qy        405 SEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTV 464
                      |  :   |  :: |:||:      |: |         | :|  |  ||  |
Db        859 -------PVTVSAVKP--AVTTLVVKGTTGVTTLGTVT------GTVSTSLAGAGGHST- 902

Qy        465 IIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTT 524
                           | | ||  |      || : :  :  |||  |        |  |
Db        903 ---------------SASLATPITTL----GTIATLSSQVINPTAITVS------AAQT 936

Qy        525 TVTAPPGGTDSVIIREP-PNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTV- 582
              |:||  | |   |  :|   ||           | :||| |      :   | | :
Db        937 TLTAAGGLTTPTITMQPVSQPT-----------QVTLITAPSG------VEAQPVHDLP 978

Qy        583 ------TTTEYWSQSYATTTTVTAPPG----GTDTVIIREPP--NHTVTTTEYWSQSFAT 630
                    |||   |  || |   :  |    || |::   ||   |   ||      : ||
Db        979 VSILASPTTE---QPTATVTIADSGQGDVQPGTVTLVCSNPPCETHETGTT-----NTAT 1030

Qy        631 TTTVT---GPPSG----------------TDTV------IIREPPNPTVTTTEYWSQSY 664
              || |    | |                  | ||      ::|   ||   |  |  : :
Db       1031 TTVVANLGGHPQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCETHETGTTNT 1090

Qy        665 ATTTT--------ITAPPGET---------DTVLIREPPNH-----------------T 689
              |||  |         : || ||          | :     ||                 :
Db       1091 ATTATSNMAGQHGCSNPPCETHETGTTNTATTAMSSVGANHQRDARRACAAGTPAVIRIS 1150

Qy        690 VTTTEYWS----------QSYATTTTVT-----AP-------------PGETDTVLIR 719
              | |    :           |: ||:||:|     ||             ||      ::
Db       1151 VATGALEAAQGSKSQCQTRQTSATSTTMTVMATGAPCSAGPLLGPSMAREPGGRSPAFVQ 1210

Qy        720 EPP-------------------NHTVTTTEYWSQSY------------------ATTT 740
                |                   :| |:|      |                   :||
Db       1211 LAPLSSKVRLSSPSIKDLPAGRHSHAVSTAAMTRSSVGAGEPRMAPVCESLQGGSPSTTV 1270

Qy        741 TVTA-----PPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMS 795
              ||||      |  | | :   ||  |        ::  | |  |:  |    |
Db       1271 TVTALEALLCPSATVTQVCSNPPCET-------HETGTTNTATTSNAGSAQRVCSNPPCE 1323

Qy        796 SSKISTSSNDITSII------------PSFSRPHYVNSTTSDLSTFE-SSSMNTPTSISS 842
              : :  |:    |:              |   ||   : |||  :|   |     | : ||

Db       1324 THETGTTHTATTATSNGGTGQPEGGQQPPAGRPCETHQTTSTGTTMSVSVGALLPDATSS 1383

Qy        843 -----DGMLLSSTTLVTESETT-------TELICSDGKECSRLSSSSGIVTNP--DSNES 888
                   |: :::   ||    |       |: :||                ||  :::|
Db       1384 HRTVESGLEVAAAPSVTPQAGTALLAPFPTQRVCS---------------NPPCETHE- 1426

Qy        889 SIVTSTVPTASTMSDSLSSTDGISATSSD-----NVSKSGVSVTTETSVTTIQTTPNPLS 943
                | |  ||:|:: ::||        :||      :     |::|: :::||  |  : |:
Db       1427 ---TGTTHTATTVTSNMSSNQDPPPAASDQGEVESTQGDSVNITSSSAITT--TVSSTLT 1481

Qy        944 SSVTSLTQLSSIPSVSESESKVTFTSNGDNQS--GTHDSQSTSTEI-----EIVTTSSTK 996
               :||::|| : :|  |    :    | |  |         || ||  :     |::: | |
Db       1482 RAVTTVTQSTPVPGPSVPPPEELQVSPGPRQQLPPRQLLQSASTALMGESAEVLSASQTP 1541

Qy        997 VLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVA 1056
               ||  |  ::  | ||:: :|       :: |        ||                | || :
Db       1542 ELPAAVDLSS--TGEPSSGQE-------SAGSAVVATVVVQP------------PPPTQS 1580

Qy       1057 TSTLASASEE---DNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSS 1113
                    |  :|    : ::|:     | | |    |   :| : |  |   :  || : |: :
Db       1581 EVDQLSLPQELMAEAQAGTTTLMVTGLTP---EELAVTAAAEAAAQAAATEEAQALAIQA 1637

Qy       1114 GTDV-------TTEPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTT 1166
                       | || || |   |:      | :||        :  | :  :| :   ||
Db       1638 VLQAAQQAVMGTGEPMDTSEAAATV-----TQAEL----GHLSAEGQEGQATTIPIVLTQ 1688

Qy       1167 GTSASTSANSELVTSGS--------VTGGAVASASNDQSHSTSVTNSNSIVSNTPQTT-- 1216
                  |:      :|  : :              | | : || :  ::    |  :    | |
Db       1689 QELAALVQQQQLQEAQAQQQHHHLPTEALAPADSLNDPAIESNCL--NELAGTVPSTVAL 1746

Qy       1217 LSQQVTSSSPSTNTFIA 1233
                |    | |   :|||:|
Db       1747 LPSTATESLAPSNTFVA 1763


RESULT 10
US-08-393-703-5
; Sequence 5, Application US/08393703
; Patent No. 5585239
;  GENERAL INFORMATION:
;    APPLICANT:  Lamarco, Kelly
;    APPLICANT:  Wilson, Angus
;    APPLICANT:  Herr, Winship
;    TITLE OF INVENTION:  A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
;    TITLE OF INVENTION:  HOST CELL FACTOR
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/393,703
;      FILING DATE:  24-FEB-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-57503-2/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2035 amino acids
```

```
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-393-703-5

  Query Match             6.2%;  Score 404.5;  DB 1;  Length 2035;
  Best Local Similarity   21.9%;  Pred. No. 4.9e-15;
  Matches  319;  Conservative  165;  Mismatches  506;  Indels  467;  Gaps   69;

Qy         14 IASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAV-LGWSLDGTSA-NPGDTFTLNM 71
              | |:  ::|:     :|  : ||          ||  :|  | ::  | |  ||   |  :.
Db        537 IGSSPQMSGM----AALAAAAAATQKIPPSSAPTVLSVPAGTTIVKTMAVTPG---TTTL 589

Qy         72 PCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLP 131
              |   |  :|   |                |       |    |  :: |   |  | | |
Db        590 PATVKVASSPVMV--------------SNPATRMLKTAAAQVGTSVSS---ATNTSTRP 631

Qy        132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMP 191
              |   |  :|: |  : ::  |     ||     | |::          | :    | ::
Db        632 I-ITVHKSGTVTVAQQAQVVTTVVGGVT-----KTITL-------VKSPISVPGGSALIS 678

Qy        192 SLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFS 251
              :| || ::    | :   ||   |  :| | |         |  |||
Db        679 NLGKVMSVV---QTKPVQTSAVTG-QASTGPVT------QIIQTKG-------------- 714

Qy        252 YTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLR 311
                        :|||     |   ::: |     |:  |   : ||    :||  |
Db        715 --------------PLPAG---TILKLVTSADGKPTTIITTTQASGAG----TKPTIL- 751

Qy        312 WTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDK------TKTIEILQPIP-- 363
                       ||  |: : |   :|| : |:| :  : :      | :  |  ||
Db        752 ------------GISSVSPSTTKPGTTTIIKTIPMSAIITQAGATGVTSSPGIKSPITII 799

Qy        364 TTTITTSYVGVTTSYLT-------------------KTAPIGETATVIVDVPYHTTTTVT 404
              || : ||  |     :|                   | || |:  |::  ||      ||
Db        800 TTKVMTSGTGAPAKIITAVPKIATGHGQQGVTQVVLKGAP-GQPGTILRTVPMGGVRLVT 858

Qy        405 SEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTV 464
                      | :   |  :: |:||:      |: |       | :|  |  ||  |
Db        859 -------PVTVSAVKP--AVTTLVVKGTTGVTTLGTVT------GTVSTSLAGAGGHST- 902

Qy        465 IIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTT 524
                        | | ||  |     || : :  :  |||  |       |  |
Db        903 ----------------SASLATPITTL----GTIATLSSQVINPTAITVS------AAQT 936

Qy        525 TVTAPPGGTDSVIIREP-PNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTV- 582
              |:||  | |   |  :|   ||              | :||| |      :   | | :
Db        937 TLTAAGGLTTPTITMQPVSQPT-----------QVTLITAPSG------VEAQPVHDLP 978

Qy        583 ------TTTEYWSQSYATTTTVTAPPG----GTDTVIIREPP--NHTVTTTEYWSQSFAT 630
                   |||   |  || |   :  |    || |::   ||   |   ||      : ||
Db        979 VSILASPTTE---QPTATVTIADSGQGDVQPGTVTLVCSNPPCETHETGTT-----NTAT 1030

Qy        631 TTTVT---GPPSG-----------------TDTV------IIREPPNPTVTTTEYWSQSY 664
              || |    | |                   | ||      ::|   ||   | |  : :
Db       1031 TTVVANLGGHPQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCETHETGTTNT 1090

Qy        665 ATTTT--------ITAPPGET---------DTVLIREPPNH-----------------T 689
              |||  |         : || ||          | :     ||                 :
Db       1091 ATTATSNMAGQHGCSNPPCETHETGTTNTATTAMSSVGANHQRDARRACAAGTPAVIRIS 1150

Qy        690 VTTTEYWS-----------QSYATTTTVT-----AP--------------PGETDTVLIR 719
              | |    :           |: ||:||:|     ||                ||      ::
Db       1151 VATGALEAAQGSKSQCQTRQTSATSTTMTVMATGAPCSAGPLLGPSMAREPGGRSPAFVQ 1210

Qy        720 EPP-------------------NHTVTTTEYWSQSY-------------------ATTT 740
                |                  :| |:|     |                      :||
Db       1211 LAPLSSKVRLSSPSIKDLPAGRHSHAVSTAAMTRSSVGAGEPRMAPVCESLQGGSPSTTV 1270

Qy        741 TVTA-----PPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMS 795
              ||||      |  | | :   ||  |        ::  | |  |:  |    |
Db       1271 TVTALEALLCPSATVTQVCSNPPCET-------HETGTTNTATTSNAGSAQRVCSNPPCE 1323

Qy        796 SSKISTSSNDITSII-----------PSFSRPHYVNSTTSDLSTFE-SSSMNTPTSISS 842
              : :  |:    |:              |   ||   : |||  :|   |     | : ||
Db       1324 THETGTTHTATTATSNGGTGQPEGGQQPPAGRPCETHQTTSTGTTMSVSVGALLPDATSS 1383

Qy        843 -----DGMLLSSTTLVTESETT--------TELICSDGKECSRLSSSSGIVTNP--DSNES 888
                       |: :::   ||    |       |: :||                ||  :::|
Db       1384 HRTVESGLEVAAAPSVTPQAGTALLAPFPTQRVCS----------------NPPCETHE- 1426

Qy        889 SIVTSTVPTASTMSDSLSSTDGISATSSD-----NVSKSGVSVTTETSVTTIQTTPNPLS 943
                 | |  ||:|:: ::||       :||     :     |::|: :::||  |  : |:
Db       1427 ---TGTTHTATTVTSNMSSNQDPPPAASDQGEVESTQGDSVNITSSSAITT--TVSSTLT 1481

Qy        944 SSVTSLTQLSSIPSVSESESKVTFTSNGDNQS--GTHDSQSTSTEI-----EIVTTSSTK 996
               :||::|| : :|  |    :    | |  |         || || :     |::: | |
Db       1482 RAVTTVTQSTPVPGPSVPPPEELQVSPGPRQQLPPRQLLQSASTALMGESAEVLSASQTP 1541

Qy        997 VLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVA 1056
               ||  |  ::  | ||:: :|       :: |     ||              | || :
Db       1542 ELPAAVDLSS--TGEPSSGQE-------SAGSAVVATVVVQP-----------PPPTQS 1580

Qy       1057 TSTLASASEE---DNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSS 1113
                  |  :|   : ::|:     | | |   |   :| : |  |   :  || : |: :
Db       1581 EVDQLSLPQELMAEAQAGTTTLMVTGLTP---EELAVTAAAEAAAQAAATEEAQALAIQA 1637

Qy       1114 GTDV-------TTEPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTT 1166
                       | || || |   |:     | :||       :  | :  :| :   ||
Db       1638 VLQAAQQAVMGTGEPMDTSEAAATV-----TQAEL----GHLSAEGQEGQATTIPIVLTQ 1688

Qy       1167 GTSASTSANSELVTSGS--------VTGGAVASASNDQSHSTSVTNSNSIVSNTPQTT-- 1216
                  |:     :|  : :           | | : || :  ::    | :    | |
Db       1689 QELAALVQQQQLQEAQAQQQHHHLPTEALAPADSLNDPAIESNCL--NELAGTVPSTVAL 1746

Qy       1217 LSQQVTSSSPSTNTFIA 1233
                |    | |  . :|||:|
Db       1747 LPSTATESLAPSNTFVA 1763


RESULT 11
PCT-US93-11721-5
; Sequence 5, Application PC/TUS9311721
;  GENERAL INFORMATION:
;    APPLICANT:  Lamarco, Kelly
;    APPLICANT:  Wilson, Angus
;    APPLICANT:  Herr, Winship
;    TITLE OF INVENTION:  A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:
;    TITLE OF INVENTION:  HOST CELL FACTOR
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/11721
;      FILING DATE:  03-DEC-1993
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  FP-57503-1/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299
;  INFORMATION FOR SEQ ID NO:  5:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2035 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
PCT-US93-11721-5

  Query Match             6.2%;  Score 404.5;  DB 5;  Length 2035;
  Best Local Similarity   21.9%;  Pred. No. 4.9e-15;
  Matches  319;  Conservative  165;  Mismatches  506;  Indels  467;  Gaps   69;

Qy         14 IASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAV-LGWSLDGTSA-NPGDTFTLNM 71
              | |:  ::|:     :|  : ||          ||  :|  | ::  | |  ||   |  :
Db        537 IGSSPQMSGM----AALAAAAAATQKIPPSSAPTVLSVPAGTTIVKTMAVTPG---TTTL 589

Qy         72 PCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLP 131
              |   | :|   |                |       |    | :: |   |  | | |
Db        590 PATVKVASSPVMV---------------SNPATRMLKTAAAQVGTSVSS---ATNTSTRP 631

Qy        132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMP 191
              |   |  :|: |  : ::  |     ||     | |::          | :    | ::
Db        632 I-ITVHKSGTVTVAQQAQVVTTVVGGVT-----KTITL-------VKSPISVPGGSALIS 678

Qy        192 SLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFS 251
              :| || ::    | :   ||    |  :| | |         |  |||
Db        679 NLGKVMSVV---QTKPVQTSAVTG-QASTGPVT------QIIQTKG-------------- 714

Qy        252 YTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLR 311
                              :|||    |   ::: |    |:  |   : ||    :||  |
Db        715 ---------------PLPAG---TILKLVTSADGKPTTIITTTQASGAG----TKPTIL- 751

Qy        312 WTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDK------TKTIEILQPIP-- 363
                           ||  |: : |   :|| : |:| :  : :       | :  |  ||
Db        752 ------------GISSVSPSTTKPGTTTIIKTIPMSAIITQAGATGVTSSPGIKSPITII 799

Qy        364 TTTITTSYVGVTTSYLT-------------------KTAPIGETATVIVDVPYHTTTTVT 404
              || : ||  |     :|                  | || |:  |::  ||       ||
Db        800 TTKVMTSGTGAPAKIITAVPKIATGHGQQGVTQVVLKGAP-GQPGTILRTVPMGGVRLVT 858

Qy        405 SEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTV 464
                      | :   |  :: |:||:      |: |         | :|  |  ||  |
Db        859 -------PVTVSAVKP--AVTTLVVKGTTGVTTLGTVT------GTVSTSLAGAGGHST- 902

Qy        465 IIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTT 524
                          | | ||  |      || : :  :  |||  |         |  |
Db        903 ---------------SASLATPITTL----GTIATLSSQVINPTAITVS------AAQT 936

Qy        525 TVTAPPGGTDSVIIREP-PNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTV- 582
              |:||  | |   | :|   ||              | :||| |       :   | | :
Db        937 TLTAAGGLTTPTITMQPVSQPT------------QVTLITAPSG------VEAQPVHDLP 978

Qy        583 ------TTTEYWSQSYATTTTVTAPPG----GTDTVIIREPP--NHTVTTTEYWSQSFAT 630
                    |||   |  || |   :  |    || |::   ||   |   ||     : ||
Db        979 VSILASPTTE---QPTATVTIADSGQGDVQPGTVTLVCSNPPCETHETGTT-----NTAT 1030

Qy        631 TTTVT---GPPSG----------------TDTV------IIREPPNPTVTTTEYWSQSY 664
              || |    | |                 | ||      ::|   ||   | |  : :
Db       1031 TTVVANLGGHPQPTQVQFVCDRQEAAASLVTSTVGQQNGSVVRVCSNPPCETHETGTTNT 1090

Qy        665 ATTTT--------ITAPPGET---------DTVLIREPPNH----------------T 689
              ||| |          : || ||         | :     ||                :
Db       1091 ATTATSNMAGQHGCSNPPCETHETGTTNTATTAMSSVGANHQRDARRACAAGTPAVIRIS 1150

Qy        690 VTTTEYWS----------QSYATTTTVT-----AP--------------PGETDTVLIR 719
              | |    :           |: ||:||:|     ||              ||      ::
Db       1151 VATGALEAAQGSKSQCQTRQTSATSTTMTVMATGAPCSAGPLLGPSMAREPGGRSPAFVQ 1210

Qy        720 EPP-------------------NHTVTTTEYWSQSY-----------------ATTT 740
                |                   :| |:|     |                   :||
Db       1211 LAPLSSKVRLSSPSIKDLPAGRHSHAVSTAAMTRSSVGAGEPRMAPVCESLQGGSPSTTV 1270

Qy        741 TVTA-----PPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMS 795
                          |||:      |  | | :   ||  |        ::  | |  |:  |    |
Db       1271 TVTALEALLCPSATVTQVCSNPPCET-------HETGTTNTATTSNAGSAQRVCSNPPCE 1323

Qy        796 SSKISTSSNDITSII------------PSFSRPHYVNSTTSDLSTFE-SSSMNTPTSISS 842
              : :  |:    |:              |   ||   : |||  :|   |     | : ||
Db       1324 THETGTTHTATTATSNGGTGQPEGGQQPPAGRPCETHQTTSTGTTMSVSVGALLPDATSS 1383

Qy        843 -----DGMLLSSTTLVTESETT-------TELICSDGKECSRLSSSSGIVTNP--DSNES 888
                    |: :::   ||    |       |: :||                  ||  :::|
Db       1384 HRTVESGLEVAAAPSVTPQAGTALLAPFPTQRVCS---------------NPPCETHE- 1426

Qy        889 SIVTSTVPTASTMSDSLSSTDGISATSSD-----NVSKSGVSVTTETSVTTIQTTPNPLS 943
                | |  ||:|:: ::||       :||     :     |::|: :::||  |  : |:
Db       1427 ---TGTTHTATTVTSNMSSNQDPPPAASDQGEVESTQGDSVNITSSSAITT--TVSSTLT 1481

Qy        944 SSVTSLTQLSSIPSVSESESKVTFTSNGDNQS--GTHDSQSTSTEI-----EIVTTSSTK 996
               :||::|| : :|  |    :    | |  |         || || :     |::: | |
Db       1482 RAVTTVTQSTPVPGPSVPPPEELQVSPGPRQQLPPRQLLQSASTALMGESAEVLSASQTP 1541

Qy        997 VLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVA 1056
               ||  |  ::  | ||:: :|      :: |         ||             | || :
Db       1542 ELPAAVDLSS--TGEPSSGQE-------SAGSAVVATVVVQP------------PPPTQS 1580

Qy       1057 TSTLASASEE---DNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSS 1113
                  | :|   : ::|:     | | |   |   :| : |  |  :  || : |: :
Db       1581 EVDQLSLPQELMAEAQAGTTTLMVTGLTP---EELAVTAAAEAAAQAAATEEAQALAIQA 1637

Qy       1114 GTDV-------TTEPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTT 1166
                         | || || |   |:      | :||       :  | :  :| :   ||
Db       1638 VLQAAQQAVMGTGEPMDTSEAAATV-----TQAEL----GHLSAEGQEGQATTIPIVLTQ 1688

Qy       1167 GTSASTSANSELVTSGS--------VTGGAVASASNDQSHSTSVTNSNSIVSNTPQTT-- 1216
                 |:     :|  : :           | | : || :  ::    | :    | |
Db       1689 QELAALVQQQQLQEAQAQQQHHHLPTEALAPADSLNDPAIESNCL--NELAGTVPSTVAL 1746

Qy       1217 LSQQVTSSSPSTNTFIA 1233
                 |    | |   :|||:|
Db       1747 LPSTATESLAPSNTFVA 1763


RESULT 12
US-09-071-035-258
; Sequence 258, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
```

```
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  258:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1638 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-258

  Query Match             6.0%;  Score 392;  DB 4;  Length 1638;
  Best Local Similarity   22.4%;  Pred. No. 1.9e-14;
  Matches  311;  Conservative  149;  Mismatches  603;  Indels  326;  Gaps   60;

Qy         13 SIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVL-----GWSLDGTSANPG-DT  66
Db        351 SEVTAPTITGV-------TGNSTAGYEVKGTA--DANATVEIRNAGGTVIGTGTADGTGA  401

Qy         67 FTLNMPC-VFKYTTSQTSVDLTADGVKY--ATCQFYSGEEFTTFSTLT-----CTVNDAL  118
Db        402 FTVTVPAGEAGANETLTAVAKNASGTEXTPTTFQTPADEATVTAPTITGVTGNSTAGYEV  461

Qy        119 KSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV  178
Db        462 KGTADANATVEIR---NAGGTVIGTGTADG----TGAFTVTVPAGEAGAN-----ETLTA  509

Qy        179 DPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGL  238
Db        510 -------VAKNASGTESTPTTFQTPADEATVTAPTITGVTGNSTAGYE--------VKGT  554

Qy        239 NDWNYPVS--SESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDY  296
Db        555 ADANATVEIRNAGGAVIGTGTADGTGAFTVTIPAG----------EAGANETLTAVAKN-  603

Qy        297 TCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTI  356
Db        604 ---ASGTESTPTTFQ------TPADPNTPVATPIVETVTGSTTKGYEVKGTAEVG--TTI  652

Qy        357 EILQPIPTTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTT  416
Db        653 EV-RDAAGTVLGTATTGTDGKY-TVTLDSG-TATA-----NQTLSVVAKNASGTESQPAT  704

Qy        417 RTNPTD----SIDTV---------VVQVPLPNPTVSTTEYWSQSFATTT-------TVTA  456
Db        705 ATTPADVTAPTVDNITGNSGSGYEITGTADPNTTIEVRDPSGAVIGTGTSDANGDFTVTL  764

Qy        457 P-----PGGTDTVIIREPPNHTVTTTEYWSQSFATTT-----------------TVTAPP  494
Db        765 PTGTTNPGDTLTVIGKDNAGNESQPTEVLVPADATVTAPTVTGVTGNSVAGYQVTGTADP  824

Qy        495 GGTDSVIIREPP-NPTVTTTEYWSQSFATTTTV-TAPPGGTDSVIIREPPNPTVTTTEYW  552
Db        825 NAT--IEIRDADGNVIATGTADGTGSFAVNLPAGTANANETLTALAKDPAGNTSTPTTFQ  882

Qy        553 SQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPG---GTDTVI  609
Db        883 TPA----DEVVAPP-SVDKVTGNTTQGYQVTGT----AELGTTIEVRATDGTVLGTAT--  931

Qy        610 IREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVII--------REPPNPTVTTTEYWS  661
Db        932 -------TGPTGQY------TVTLASGKATAKQTVNVVAKNDTGLESQPTTAMTPAD---  975

Qy        662 QSYATTTTI--------------TAPPGETDTVLIREPPNHTVTTTEYWSQSYATT--T  704
Db        976 ---VTTPTIGDITGDSTTGYEITGTADP--NTTIEVRNPDGTIIGTTTTDDQGNFTVDLP 1030

Qy        705 TVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTT--TVTAPPGGTDT---VIIREPPN  759
Db       1031 AGAANPGDTLTVVGKDGDGNESQPTEVTVPEDATVAAPTVTTVTGTTATGYQVTGTAEPN 1090

Qy        760 PTVTTTEYWSQSFATTT-------TVTAPPGGTDTVIIYESMSSSKISTSSN-------- 804
Db       1091 VTIEIHNEAGLVIATGTTDGAGAFTITLPTGTATANEALTAIAKDAAGKESNPTAFKTPA 1150

Qy        805 --DITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI-SSDGMLLSSTTLVTESETTTE 861
Db       1151 DPDAPVATPTVDK--ITGSTTNGYQVVGAAEVGTTVEVRDADGTVLGMATTGTDGKYTVT 1208

Qy        862 LICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPT--ASTMSDSLSSTDGISATSSDNV 919
Db       1209 L--EPGK-ASANETITVVAKNATGKESQPATATTPVDLATPTIDSITGN----------- 1254

Qy        920 SKSGVSVT-TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTH 978
Db       1255 SSKGYEITGTAEPKTTIDVR---------------------------------DADGTI 1280

Qy        979 DSQSTSTEIEIVTTSSTKVLPPVV----SSNTDLTSEPTNTREQPTTLSTTSNSITEDIT 1034
Db       1281 IAATTANE----TGQYTVTLPAGVVTPGETITIISKDGAGNESQPATAVIPADVVLAAPT 1336

Qy       1035 TSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTST 1094
Db       1337 ITKVEGNKANGYTVTGTADPNVTVQPYNSSEQLLASGNTTTGGTF---SVHIAAGLATEK 1393

Qy       1095 EIEATTTSPTE---APSPAVSSGTDVTTEPTDTREQPTTLST--TSKTNSELVATT---- 1145
Db       1394 ETLTALTTDTQGNVSPKTTFMTPADITGEPEIKIAAPTVSSVLGTSKAGYLIKGTAEPNR 1453

Qy       1146 ----------------------------QATNENGGKSPSTDLTSSLTTGTSASTSA 1174
Db       1454 IIQISNRLLRSVIAVGATDAEGNFAIQLTAGQATAQQSLLATATDGAGHYSTATTFMTPA 1513

Qy       1175 NSELVTSGS-VTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIA 1233
Db       1514 DPTNPGGGNGNTGGNNGNTGGNTGNNGATGGNNGNGSN----------TGSNPNGGSGLG 1563

Qy       1234 STYDGSGSI 1242
Db       1564 TTGSGLGSL 1572


RESULT 13
US-09-071-035-262
; Sequence 262, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  262:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1638 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-262

  Query Match              6.0%;  Score 392;  DB 4;  Length 1638;
  Best Local Similarity   22.4%;  Pred. No. 1.9e-14;
  Matches  311;  Conservative  149;  Mismatches  603;  Indels  326;  Gaps   60;

Qy         13 SIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVL-----GWSLDGTSANPG-DT 66
              |  :| |||||       | :: | |  ||      || :     | :: ||    |
Db        351 SEVTAPTITGV-------TGNSTAGYEVKGTA--DANATVEIRNAGGTVIGTGTADGTGA 401

Qy         67 FTLNMPC-VFKYTTSQTSVDLTADGVKY--ATCQFYSGEEFTTFSTLT-----CTVNDAL 118
              ||: :|        : |:|   | | :    | |  : |   |  |:|      |    :
Db        402 FTVTVPAGEAGANETLTAVAKNASGTEXTPTTFQTPADEATVTAPTITGVTGNSTAGYEV 461

Qy        119 KSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV 178
              | :  |  || :   | |||   |   |      |  |||   |:   :     |  |
Db        462 KGTADANATVEIR---NAGGTVIGTGTADG----TGAFTVTVPAGEAGAN-----ETLTA 509

Qy        179 DPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGL 238
                        ::       | |  |  |   |: |:   : |     :          ||
Db        510 -------VAKNASGTESTPTTFQTPADEATVTAPTITGVTGNSTAGYE--------VKGT 554

Qy        239 NDWNYPVS--SESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDY 296
              | |  |    :   :   | |::|       :|||               |:   |   :
Db        555 ADANATVEIRNAGGAVIGTGTADGTGAFTVTIPAG----------EAGANETLTAVAKN- 603

Qy        297 TCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTI 356
                |  :| | | :       : |  |  |      ||| |||    :     |    ||
Db        604 ---ASGTESTPTTFQ------TPADPNTPVATPIVETVTGSTTKGYEVKGTAEVG--TTI 652

Qy        357 EILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTT 416
              |: :    | : |:  |    | | |   | |||         | : |    :|| :   |
Db        653 EV-RDAAGTVLGTATTGTDGKY-TVTLDSG-TATA-----NQTLSVVAKNASGTESQPAT 704

Qy        417 RTNPTD----SIDTV---------VVQVPLPNPTVSTTEYWSQSFATTT-------TVTA 456
              | | |    ::| :         :     || |:   :       | |       |||
Db        705 ATTPADVTAPTVDNITGNSGSGYEITGTADPNTTIEVRDPSGAVIGTGTSDANGDFTVTL 764

Qy        457 P-----PGGTDTVIIREPPNHTVTTTEYWSQSFATTT----------------TVTAPP 494
              |     || | ||| ::   :     ||    : || |                | || |
Db        765 PTGTTNPGDTLTVIGKDNAGNESQPTEVLVPADATVTAPTVTGVTGNSVAGYQVTGTADP 824

Qy        495 GGTDSVIIREPP-NPTVTTTEYWSQSFATTTTV-TAPPGGTDSVIIREPPNPTVTTTEYW 552
                |  : ||:   |   | |   : |||      ||    | : : ::|   | | | :
Db        825 NAT--IEIRDADGNVIATGTADGTGSFAVNLPAGTANANETLTALAKDPAGNTSTPTTFQ 882

Qy        553 SQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPG---GTDTVI 609
              : :      | |||   | |       : || |        ||  | |  |   || |
Db        883 TPA----DEVVAPP-SVDKVTGNTTQGYQVTGT----AELGTTIEVRATDGTVLGTAT-- 931

Qy        610 IREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVII--------REPPNPTVTTTEYWS 661
                  | | :|      | |  :|  :   || :           |   :|  :
Db        932 -------TGPTGQY------TVTLASGKATAKQTVNVVAKNDTGLESQPTTAMTPAD--- 975

Qy        662 QSYATTTTI--------------TAPPGETDTVLIREPPNHTVTTTEYWSQSYATT--T 704
                || ||               || |    |: :| |    : ||    |   |
Db        976 ---VTTPTIGDITGDSTTGYEITGTADP--NTTIEVRNPDGTIIGTTTTDDQGNFTVDLP 1030

Qy        705 TVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTT--TVTAPPGGTDT---VIIREPPN 759
                | ||:| ||: ::   :    ||      ||    |||   | | |   |      ||
Db       1031 AGAANPGDTLTVVGKDGDGNESQPTEVTVPEDATVAAPTVTTVTGTTATGYQVTGTAEPN 1090

Qy        760 PTVTTTEYWSQSFATTT-------TVTAPPGGTDTVIIYESMSSSKISTSSN-------- 804
                |:          || |       |:| | |         :::       ||
Db       1091 VTIEIHNEAGLVIATGTTDGAGAFTITLPTGTATANEALTAIAKDAAGKESNPTAFKTPA 1150

Qy        805 --DITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI-SSDGMLLSSTTLVTESETTTE 861

                |     |:  :     |||:      :: : |   :  :|| :|   |  |: : |
Db       1151 DPDAPVATPTVDK--ITGSTTNGYQVVGAAEVGTTVEVRDADGTVLGMATTGTDGKYTVT 1208

Qy        862 LICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPT--ASTMSDSLSSTDGISATSSDNV 919
              |    || |   : : :  |    ||   |:| |   |:   ||::
Db       1209 L--EPGK-ASANETITVVAKNATGKESQPATATTPVDLATPTIDSITGN---------- 1254

Qy        920 SKSGVSVT-TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTH 978
              |  | :| |    |||                                      :  ||
Db       1255 SSKGYEITGTAEPKTTIDVR--------------------------------DADGTI 1280

Qy        979 DSQSTSTEIEIVTTSSTKVLPPVV----SSNTDLTSEPTNTREQPTTLSTTSNSITEDIT 1034
               : :|: |    |   |  || |     : | :: :      || |    :: :    |
Db       1281 IAATTANE----TGQYTVTLPAGVVTPGETITIISKDGAGNESQPATAVIPADVVLAAPT 1336

Qy       1035 TSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTST 1094
               ::  |:  :  : |       |    ::||:   ||: :  |    |:   :|| |
Db       1337 ITKVEGNKANGYTVTGTADPNVTVQFYNSSEQLLASGNTTTGGTF---SVHIAAGLATEK 1393

Qy       1095 EIEATTTSPTE---APSPAVSSGTDVTTEPTDTREQPTTLST--TSKTNSELVATT---- 1145
              |     |: |:   :|     :  |:| ||      ||  |   |||    :  |
Db       1394 ETLTALTTDTQGNVSPKTTFMTPADITGEPEIKIAAPTVSSVLGTSKAGYLIKGTAEPNR 1453

Qy       1146 -------------------------------QATNENGGKSPSTDLTSSLTTGTSASTSA 1174
                                            ||| :    : :||     :| |:  | |
Db       1454 IIQISNRLLRSVIAVGATDAEGNFAIQLTAGQATAQQSLLATATDGAGHYSTATTFMTPA 1513

Qy       1175 NSELVTSGS-VTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIA 1233
              :      |:  |||   :   :  :: :   :|   ||          | |:|:  : :
Db       1514 DPTNPGGGNGNTGGNNGNTGGNTGNNGATGGNNGNGSN----------TGSNPNGGSGLG 1563

Qy       1234 STYDGSGSI 1242
              :|  | ||:
Db       1564 TTGSGLGSL 1572


RESULT 14
US-09-071-035-266
; Sequence 266, Application US/09071035
; Patent No. 6448043
;  GENERAL INFORMATION:
;    APPLICANT:  Gil H. Choi
;    TITLE OF INVENTION:  Enterococcus faecalis Polynucleotides and Polypeptides
;    NUMBER OF SEQUENCES:  496
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,035
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  A. Anders Brookes
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB369P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  266:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1638 amino acids
```

```
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-071-035-266

  Query Match             6.0%;  Score 392;  DB 4;  Length 1638;
  Best Local Similarity   22.4%;  Pred. No. 1.9e-14;
  Matches  311;  Conservative  149;  Mismatches  603;  Indels  326;  Gaps   60;

Qy         13 SIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVL-----GWSLDGTSANPG-DT  66
              |  :| |||||       | :: | |  ||      || :     | :: ||    |
Db        351 SEVTAPTITGV-------TGNSTAGYEVKGTA--DANATVEIRNAGGTVIGTGTADGTGA 401

Qy         67 FTLNMPC-VFKYTTSQTSVDLTADGVKY--ATCQFYSGEEFTTFSTLT-----CTVNDAL 118
              ||: :|       : |:|   | | :    | |  : |   |  |:|      |    :
Db        402 FTVTVPAGEAGANETLTAVAKNASGTEXTPTTFQTPADEATVTAPTITGVTGNSTAGYEV 461

Qy        119 KSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV 178
              | :  |  || :    | |||   |   |      |  |||   |:   :     |  |
Db        462 KGTADANATVEIR---NAGGTVIGTGTADG----TGAFTVTVPAGEAGAN-----ETLTA 509

Qy        179 DPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGL 238
                    ::          | |  | |   |: |:   : |     :          ||
Db        510 -------VAKNASGTESTPTTFQTPADEATVTAPTITGVTGNSTAGYE--------VKGT 554

Qy        239 NDWNYPVS--SESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDY 296
               | |  |   :   :  | |::|       :|||              |:   |   :
Db        555 ADANATVEIRNAGGAVIGTGTADGTGAFTVTIPAG----------EAGANETLTAVAKN- 603

Qy        297 TCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTI 356
                  |  :| | | :      : |  |  |      ||| |||    :      |    ||
Db        604 ---ASGTESTPTTFQ------TPADPNTPVATPIVETVTGSTTKGYEVKGTAEVG--TTI 652

Qy        357 EILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTT 416
              |: :    | : |:  |    | | |   | |||         | : |    :|| :   |
Db        653 EV-RDAAGTVLGTATTGTDGKY-TVTLDSG-TATA-----NQTLSVVAKNASGTESQPAT 704

Qy        417 RTNPTD----SIDTV---------VVQVPLPNPTVSTTEYWSQSFATTT-------TVTA 456
               | |  |   ::| :          :     || |:   :      | |        |||
Db        705 ATTPADVTAPTVDNITGNSGSGYEITGTADPNTTIEVRDPSGAVIGTGTSDANGDFTVTL 764

Qy        457 P-----PGGTDTVIIREPPNHTVTTTEYWSQSFATTT----------------TVTAPP 494
              |     || | ||| ::   :    ||    : || |                | || |
Db        765 PTGTTNPGDTLTVIGKDNAGNESQPTEVLVPADATVTAPTVTGVTGNSVAGYQVTGTADP 824

Qy        495 GGTDSVIIREPP-NPTVTTTEYWSQSFATTTTV-TAPPGGTDSVIIREPPNPTVTTTEYW 552
                |  : ||:   |   | |   : |||      ||    | : : ::|   | | | :
Db        825 NAT--IEIRDADGNVIATGTADGTGSFAVNLPAGTANANETLTALAKDPAGNTSTPTTFQ 882

Qy        553 SQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPG---GTDTVI 609
              : :      | |||   | |       : || |          ||  | |  |   || |
Db        883 TPA----DEVVAPP-SVDKVTGNTTQGYQVTGT----AELGTTIEVRATDGTVLGTAT-- 931

Qy        610 IREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVII--------REPPNPTVTTTEYWS 661
                  |  | :|       | |  :|  :   || :           |   :|  :
Db        932 -------TGPTGQY------TVTLASGKATAKQTVNVVAKNDTGLESQPTTAMTPAD--- 975

Qy        662 QSYATTTTI--------------TAPPGETDTVLIREPPNHTVTTTEYWSQSYATT--T 704
                || ||               || |    |: :| |    : ||    |   |
Db        976 ---VTTPTIGDITGDSTTGYEITGTADP--NTTIEVRNPDGTIIGTTTTDDQGNFTVDLP 1030

Qy        705 TVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTT--TVTAPPGGTDT---VIIREPPN 759
                | ||:| ||: ::   :    ||      ||     |||    | | |   |      ||
Db       1031 AGAANPGDTLTVVGKDGDGNESQPTEVTVPEDATVAAPTVTTVTGTTATGYQVTGTAEPN 1090

Qy        760 PTVTTTEYWSQSFATTT-------TVTAPPGGTDTVIIYESMSSSKISTSSN-------- 804
               |:          || |       |:| | |          :::        ||
Db       1091 VTIEIHNEAGLVIATGTTDGAGAFTITLPTGTATANEALTAIAKDAAGKESNPTAFKTPA 1150

Qy        805 --DITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI-SSDGMLLSSTTLVTESETTTE 861
                |     |:  :     |||:      ::: : |   :   :|| `:|   |  |: : |
Db       1151 DPDAPVATPTVDK--ITGSTTNGYQVVGAAEVGTTVEVRDADGTVLGMATTGTDGKYTVT 1208

Qy        862 LICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPT--ASTMSDSLSSTDGISATSSDNV 919
              |    ||  |   : : :  |    ||   |:| |   |:   ||::
Db       1209 L--EPGK-ASANETITVVAKNATGKESQPATATTPVDLATPTIDSITGN---------- 1254

Qy        920 SKSGVSVT-TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTH 978
              |  |  :| |    |||                                    :  ||
Db       1255 SSKGYEITGTAEPKTTIDVR--------------------------------DADGTI 1280

Qy        979 DSQSTSTEIEIVTTSSTKVLPPVV----SSNTDLTSEPTNTREQPTTLSTTSNSITEDIT 1034
               : :|: |    |   |  ||  |     : | :: :      || |    :: :    |
Db       1281 IAATTANE----TGQYTVTLPAGVVTPGETITIISKDGAGNESQPATAVIPADVVLAAPT 1336

Qy       1035 TSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTST 1094
               ::  |:  :  : |       |    ::||:   ||:  :  |    |:   :|| |
Db       1337 ITKVEGNKANGYTVTGTADPNVTVQFYNSSEQLLASGNTTTGGTF---SVHIAAGLATEK 1393

Qy       1095 EIEATTTSPTE---APSPAVSSGTDVTTEPTDTREQPTTLST--TSKTNSELVATT---- 1145
              |     |: |:   :|     :  |:| ||       ||  |   |||    :  |
Db       1394 ETLTALTTDTQGNVSPKTTFMTPADITGEPEIKIAAPTVSSVLGTSKAGYLIKGTAEPNR 1453

Qy       1146 -------------------------------QATNENGGKSPSTDLTSSLTTGTSASTSA 1174
                                             ||| :    : :||     :| |:  | |
Db       1454 IIQISNRLLRSVIAVGATDAEGNFAIQLTAGQATAQQSLLATATDGAGHYSTATTFMTPA 1513

Qy       1175 NSELVTSGS-VTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIA 1233
              :      |:  |||   :   :  :: :   :|   ||          | |:|:  : :
Db       1514 DPTNPGGGNGNTGGNNGNTGGNTGNNGATGGNNGNGSN----------TGSNPNGGSGLG 1563

Qy       1234 STYDGSGSI 1242
              :|  | ||:
Db       1564 TTGSGLGSL 1572


RESULT 15
US-07-757-022B-84
; Sequence 84, Application US/07757022B
; Patent No. 6433142
;  GENERAL INFORMATION:
;    APPLICANT:  Gesner, Thomas G.
;    APPLICANT:  Clark, Stephen C.
;    APPLICANT:  Turner, Katherine
;    APPLICANT:  Hewick, Rodney M.
;    TITLE OF INVENTION:  Megakaryocyte Stimulating Factors
;    NUMBER OF SEQUENCES:  143
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/757,022B
;      FILING DATE:  19910910
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/643,502
;      FILING DATE:  18-JAN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/546,114
;      FILING DATE:  29-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/457,196
;      FILING DATE:  29-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/390,901
;      FILING DATE:  08-AUG-1989
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cserr, Luann
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  GI 5190
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)876-1170
;      TELEFAX:  (617)876-5851
;  INFORMATION FOR SEQ ID NO:  84:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1022 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-757-022B-84

  Query Match             5.6%;  Score 364.5;  DB 4;  Length 1022;
  Best Local Similarity   23.0%;  Pred. No. 3.8e-13;
  Matches  236;  Conservative  109;  Mismatches  448;  Indels  235;  Gaps   45;

Qy        252 YTKTCTSNGIQIKYQN----VPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQ--- 304
              |   |    ::   :|     |    | :|   |  |   : :   :  |   :::
Db         73 YESFCAETAVKDNKKNRTKKKPTPKPPVVDEAGSGLDNGDFKVTTPDTSTTQHNKVSTSP 132

Qy        305 ----SKPFTLRWTGYKNSDAGSNGIVIVATTRTV-TDSTTAVTTLPFNPSVDKTKTIEIL 359
                 :||   | :   |||      : |    || |  ||           :|| : :
Db        133 KITTAKPINPRPSLPPNSDTSKETSLTVNKETTVETKETTTTNKQTSTDGKEKTTSAKET 192

Qy        360 QPIPTTTITTSYVGVTTSYLTKTAPIGETAT--VIVDVPYHTTTTVTSEWTGTI----TT 413
              | |  |  :   :  |:  | |  |  || |    :  |   | |   |   |     |
Db        193 QSIEKT--SAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTP 250

Qy        414 TTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAP-----PGGTDTVIIRE 468
              || :: ||        : | |  | |      :   |||   ||     |  | |   :|
Db        251 TTIKSAPT------TPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTPKEPAPTTT---KE 301

Qy        469 PPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTA 528
              |   |  :     :  | ||     |          |  || ||     :  | ||   |
Db        302 PAPTTTKSAPTTPKEPAPTTPKKPAPTTPKEPAPTTPKEPTPTT----PKEPAPTTKEPA 357

Qy        529 P------------------------------PGGTDSVIIREPPNPTVTTTE---YWSQ 554
              |                              |  |       |  |  |||:     ::
Db        358 PTTPKEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTTKEPSPTTPKEPAPTTTKSAPTTTK 417

Qy        555 SYATTTTVTAP--PGGTDSVIIREP----PNHTVTTTEYWSQSYATTTTVTAP-----PG 603
                | ||| :||  |        :||    |     ||    :   ||    ||     |
Db        418 EPAPTTTKSAPTTPKEPSPTTTKEPAPTTPKEPAPTTP--KKPAPTTPKEPAPTTPKEPA 475

Qy        604 GTDT-----VIIREP----PNHTVTT-----TEYWSQSFATTTTVTGPPSGTDTVIIREP 649
              | |        :||    |  |  |     |    :  | ||     |:  : :    |
Db        476 PTTTKKPAPTAPKEPAPTTPKETAPTTPKKLTPTTPEKLAPTTPEKPAPTTPEELAPTTP 535

Qy        650 PNPTVTTTEYWSQSYATTTTITAP--PGETDTVLIREP----PNHTVTTTEYWSQSYATT 703
              || || |   :   ||    ||  | |       :||    |     ||    ::  ||
Db        536 EEPTPTTPE---EPAPTTPKAAAPNTPKEPAPTTPKEPAPTTPKEPAPTTP--KETAPTT 590

Qy        704 TTVTAP--------------PGETDTVLIREP-----------PNHTVTTTEYWSQSY 736
                |||               | |       :||           |  |  ||    :
Db        591 PKGTAPTTLKEPAPTTPKKPAPKELAPTTTKEPTSTTSDKPAPTTPKGTAPTTP--KEPA 648

Qy        737 ATTTTVTAP--PGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESM 794
               ||    ||  | ||    ::||  || |  :   :  | |||     | | |     :
Db        649 PTTPKEPAPTTPKGTAPTTLKEPA-PT-TPKKPAPKELAPTTT----KGPTSTTSDKPAP 702

Qy        795 SSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVT 854
              :: | :  :   |   |: : |     || :     :| ::|||              |
Db        703 TTPKETAPT---TPKEPAPTTPKKPAPTTPETPPPTTSEVSTPT--------------T 744

Qy        855 ESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISAT 914
              | ||                |  :||  ||:   |  ||   : :|     |:  |
Db        745 TKEPTT-----------------IHKSPD--ESTPELSAEPTPKALENS-PKEPGVPTT 783

Qy        915 SSDNVSKSGVSVTTETSVT--TIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGD 972
              :   :|  :: | :   |   ::|||   :::   :|: ::  :   :|||:| |:

Db        784 KTPAATKPEMTTTAKDKTTERDLRTTPE-TTTAAPKMTKETATTTEKTTESKITATTTQV 842

Qy        973 NQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSS-NTDLTSEPTNTREQPTTLSTTSNSITE 1031
                : | |    :|  :| |  :| : | | ::  |  |:|  |  |:
Db        843 TSTTTQD----TTPFKITTLKTTTLAPKVTTTKKTITTTEIMNKPEE------------ 885

Qy       1032 DITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLT 1091
                |::|  |   |: :| | |   |     | ::   :   ::    ||       |
Db        886 ---TAKPK-DRATNSKATTPKPQKPT----KAPKKPTSTKKPKTMPRVRKPK-------T 930

Qy       1092 TSTEIEATTTSPTEAPSPAVSSG-TDVTTEPTDTREQPTTLSTTSKTNSELVATTQATNE 1150
              | |  : |:| |   |:  ::      || |  |             ||:||     : :
Db        931 TPTPRKMTSTMPELNPTSRIAEAMLQTTTRPNQT-----------PNSKLVEVNPKSED 978

Qy       1151 NGGKSPST 1158
                ||    |
Db        979 AGGAEGET 986


Search completed: June 11, 2003, 17:12:35
Job time : 30 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 17:11:32 ; Search time 53 Seconds
(without alignments)
4898.479 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQQFTLLFLYLSIASAKTI.....SIIQHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4078.5	62.8	1047	3	O93865	O93865 candida alb
2	3716.5	57.2	1523	3	Q9HFX4	Q9hfx4 candida alb
3	2791.5	43.0	1443	3	Q9Y8F2	Q9y8f2 candida alb
4	2673.5	41.2	1270	3	Q9Y743	Q9y743 candida alb
5	2576.5	39.7	2297	3	Q9HGK6	Q9hgk6 candida alb
6	1986.5	30.6	468	3	Q9URQ0	Q9urq0 candida alb
7	1823	28.1	469	3	Q9URP8	Q9urp8 candida alb
8	1780.5	27.4	468	3	Q9C471	Q9c471 candida alb
9	1471	22.6	336	3	Q9HF70	Q9hf70 candida dub
10	1191	18.3	338	3	Q9HF72	Q9hf72 candida dub
11	1176	18.1	331	3	Q9HF69	Q9hf69 candida dub
12	999.5	15.4	433	3	Q9Y8F1	Q9y8f1 candida alb
13	981	15.1	353	3	Q9HF71	Q9hf71 candida tro
14	831.5	12.8	1195	3	Q96WU8	Q96wu8 schizosacch
15	824	12.7	226	3	Q9HG16	Q9hg16 candida alb
16	698.5	10.8	1079	5	Q9N4S7	Q9n4s7 caenorhabdi
17	685	10.5	2586	5	Q9VTK8	Q9vtk8 drosophila
18	676.5	10.4	1349	4	Q8WWQ4	Q8wwq4 homo sapien
19	674.5	10.4	1220	3	Q9C0Y2	Q9c0y2 schizosacch
20	649.5	10.0	1275	5	O76602	O76602 caenorhabdi
21	646.5	10.0	2232	5	P91365	P91365 caenorhabdi
22	637.5	9.8	2271	16	Q99QY4	Q99qy4 staphylococ
23	632.5	9.7	3570	4	Q99552	Q99552 homo sapien
24	631.5	9.7	2283	2	Q8VQ99	Q8vq99 staphylococ
25	620	9.5	3072	2	Q939N5	Q939n5 streptococc
26	601	9.3	1795	5	O76894	O76894 drosophila
27	600.5	9.2	2178	2	Q9KWR3	Q9kwr3 streptococc
28	594.5	9.2	3971	3	Q96WV6	Q96wv6 schizosacch
29	588	9.1	800	3	Q8TFG4	Q8tfg4 schizosacch
30	582.5	9.0	1752	2	Q9AE52	Q9ae52 ruminococcu
31	578	8.9	1283	3	Q9USQ3	Q9usq3 schizosacch
32	577.5	8.9	957	4	O14651	O14651 homo sapien
33	577.5	8.9	1217	4	Q9UKW9	Q9ukw9 homo sapien
34	574	8.8	1283	3	Q9URU4	Q9uru4 schizosacch
35	570.5	8.8	858	3	Q9C470	Q9c470 candida alb
36	553.5	8.5	1459	5	O17084	O17084 caenorhabdi
37	553.5	8.5	3178	5	Q969D4	Q969d4 caenorhabdi
38	545.5	8.4	957	4	Q9UKN0	Q9ukn0 homo sapien
39	538.5	8.3	406	3	Q9URP9	Q9urp9 candida alb
40	538.5	8.3	1203	5	Q9N5K0	Q9n5k0 caenorhabdi
41	534	8.2	849	3	P87107	P87107 saccharomyc
42	533	8.2	407	3	O74661	O74661 candida alb
43	531.5	8.2	406	3	O74658	O74658 candida alb
44	529	8.1	407	3	O74659	O74659 candida alb
45	526	8.1	1229	5	Q94185	Q94185 caenorhabdi

ALIGNMENTS

RESULT 1
O93865
ID O93865 PRELIMINARY; PRT; 1047 AA.
AC O93865;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Agglutinin-like cell surface protein.
GN ALS8.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC10261;
RA Leng P., Lee P.R., Wishart J.A., Wu H., Brown A.J.P.;
RT "Sequence of the hypha-specific, agglutinin-like cell surface protein,
RT ALS8 from Candida albicans.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF051313; AAD02580.1; -.
SQ SEQUENCE 1047 AA; 111944 MW; C2327659AA911F2E CRC64;

Query Match 62.8%; Score 4078.5; DB 3; Length 1047;
Best Local Similarity 65.0%; Pred. No. 2.8e-163;
Matches 822; Conservative 100; Mismatches 120; Indels 223; Gaps 18;

```
Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              ||||:||| :|||:|:||||||||:|||||||||| | :|||| ||||||||||||||
Db          1 MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYHYKGPGTPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              |:|||||||||||||:|||||||||| |||||||| :|||| |||||||||:: |
Db         61 ASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGEEFMTFSTLTCTVSNTLTP 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              |||| ||||||:|||||||||| ||||||||||||||||||| | |||:|:||:| |||
Db        121 SIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTFNDGGKKISINVDFERSNVDP 180
```

```
Qy       181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
Db       181 KGYLTDSRVIPSLNKVSTLFVAPQCANGYTSGTMGFANTYGDVQIDCSNIHVGITKGLND 240

Qy       241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db       241 WNYPVSSESFSYTKTCSSNGIFITYKNVPAGYRPFVDAYISATDVNSYTLSYANEYTCAG 300

Qy       301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
Db       301 GYWQRAPFTLRWTGYRNSDAGSNGIVIVATTRTVTDSTTAVTTLPFDPNRDKTKTIEILK 360

Qy       361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
Db       361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDIPYHTTTTVTSKWTGTITSTTTHTNP 420

Qy       421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
Db       421 TDSIDTVIVQVPSPNPTVTTTEYWSQSFATTTTITGPPGNTDTVLIREPPNHTVTTTEYW 480

Qy       481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
Db       481 SESYTTTSTFTAPPGGTDSVIIKEPPNPTVTTTEYWSESYTTTSTFTAPPGGTDSVIIKE 540

Qy       541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
Db       541 PPNPTVTTTEYWSQSYTTTTTVTAPPGGTDTVLVREPPNHTVTTTEYWSQSYTTTTTVIA 600

Qy       601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
Db       601 PPGGTDTVIIREPPNH----------------------------------TVTTTEYW 624

Qy       661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
Db       625 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTITAPPGETDTVLIRE 684

Qy       721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780
Db       685 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVLIREPPNHTVTTTEYWSQSYATTTTVTA 744

Qy       781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840
Db       745 PPGGTDTVIIYDTMSSSEIS-----------SFSRPHYTNHTT----------------- 776

Qy       841 SSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900
Db       777 ------LWSTTWVIETKTITETSCEGDKGCSWVSVSTRIVTIPNNIETPMVTNTVDSTTT 830

Qy       901 MSDSLSSTDGISATSSDNVSKSGVSVTTETS-VTTIQTTPNPLSSSVTSLTQLSSIPSVS 959
Db       831 ESTS-QSPSGI-------FSESGVSVETESSTVTTAQTN----------------PSVP 865

Qy       960 ESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQP 1019
Db       866 TTESEVEFTTKGNNGNGPYESPSTH----------------VKSSMDENSEFT------ 902

Qy      1020 TTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTS 1079
Db       903 -----------------------------------TSTAAS---------------- 908

Qy      1080 LKPSMGENSGLTTSTEIE---ATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSK 1136
Db       909 -----------TSTDIENATIATTGSVEASSPIISSSADET----------TTITTTAE 946

Qy      1137 TNSELVATTQATNEN-GGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASND 1195
Db       947 STSVI---EQPTNNNGGGKAPSA------TSSPSTTTTANNDSVITG--------TTSTN 989

Qy      1196 QSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITL 1255
Db       990 QSQSQSQSN-----SDTQQTTLSQQMTSSLVSLH--MLTTFDGSGSVIQHSTWLCGLITL 1042

Qy      1256 LSLFI 1260
             |||||
Db      1043 LSLFI 1047
```

RESULT 2
Q9HFX4
ID Q9HFX4 PRELIMINARY; PRT; 1523 AA.
AC Q9HFX4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Agglutinin-like protein (Fragment).
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC5314;
RA Chen X., Chen J.-Y.;
RT "ALS4 (agglutinin-like sequence) of Candida albicans.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF272027; AAG25054.1; -.
FT NON_TER 1523 1523
SQ SEQUENCE 1523 AA; 159168 MW; 73AF3B3E442FD53C CRC64;

Query Match 57.2%; Score 3716.5; DB 3; Length 1523;
Best Local Similarity 52.0%; Pred. No. 5.7e-148;
Matches 798; Conservative 148; Mismatches 300; Indels 289; Gaps 30;

```
Qy         1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
Db         1 MLLQFLLLSLCVSVATAKVITGIFDSFNSLTWTNAASYSYRGPANPTWTAVIGWSLDGAT 60

Qy        61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
Db        61 ASAGDTFTLDMPCVFKFITDQTSIDLVADGRTYATCNLNSAEEFTTFSSVSCTVTTTMTA 120

Qy       121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db       121 DTKAIGTVTLPFSFSVGGSGSDVDLANSQCFTAGINTVTFNDGDTSISATVDFEKSTVAS 180

Qy       181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
Db       181 SDRILLSRILPSLSQAVSLFLPQECANGYTSGTMGFSTAGTGATIDCSTVHVGISNGLND 240

Qy       241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db       241 WNYPISSESFSYTKTCTSTSVLVTYQNVPAGYRPFVDAYVSATRVSSYAMRYTNIYACVG 300

Qy       301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
Db       301 AASVDDSFTHTWSGYSNSQAGSNGITIVVTTRTVTDSTTAVTTLPFNSESDKTKTIEILQ 360

Qy       361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
Db       361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy       421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
Db       421 TDSIDTVVVQVPSPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYW 480

Qy       481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
Db       481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540

Qy       541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
Db       541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTA 600

Qy       601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
Db       601 PPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYW 660
```

```
Qy        661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTV---------- 706
              |||||||||:|||| ||:|:|||||| ||||||||||||||||||
Db        661 SQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIRE 720

Qy        707 ----------------------TAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTA 744
                                   ||||| |||||||||||||||||||||||||||||||
Db        721 PPNPTVTTTEYWSQSYVTTSTITAPPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTA 780

Qy        745 PPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYE----SMSSSKIS 800
              ||||||:||||||||||||||||||:||||||||||||||||||| |    ::::::
Db        781 PPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNYTVTTTEYW 840

Qy        801 TSSNDITSIIPS---------FSRPHYVNSTTSDLSTFESSSMNTPTSIS-----SDGML 846
              : |   |: : :             |     ||::   : | |  | |:::     :| ::
Db        841 SQSYATTTTVTAPPGGTDTVIIREPPSPTVTTTE---YWSQSYATTTTVTAPPGGTDTVI 897

Qy        847 L------------------SSTTLVTESETTTELI---------CSDGKECSRLSSSSGI 879
              :                  ::|| ||      |:::       :  :  |:  :::  .
Db        898 IREPPNPTVTTTEYWSQSYATTTTVTSPPGGTDIVIIREPPNPTVTTTEYWSQSYATTTT 957

Qy        880 VTNPDSNESSIV-----TSTVPTASTMSDSLSSTDGISA------------------TSS 916
              || |     |::       || |     | | ::|  ::|                   |::
Db        958 VTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTT 1017

Qy        917 DNVSKSGVSVTTET-----------------SVTT------------------------ 934
              :  |:|  : || |                  :|||
Db       1018 EYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVI 1077

Qy        935 IQTTPNPL-------SSSVTSLTQLSSIPSVSE-----------------SES-----K 964
              |:  |||         | |  : | ::: |  ::                 |:|
Db       1078 IREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSYATTLT 1137

Qy        965 VTFTSNGDN------QSGTHDSQSTST-----------EIEIVTTSSTKVLPPVVSSNT- 1006
              :|    | |        | |:|  | ||            | ||:| |    |  |:|
Db       1138 ITAPPGGTNSVIIRVHSSTNDESSESTFSTLSVPSFSGSISIVSTVSR---PHYVNSTVT 1194

Qy       1007 ---DLTSEPTNTREQPTTLSTTSNSITE-----------DITTSQPTGDNGDNTS----- 1047
                  :|:| :         ||  ||:|             |:|:    :||  ||
Db       1195 HLPSSSSKPVDIPSSDVVPSTNDNSLTSLTGSENGETSVAISTTFCDDENGCQTSIPQGS 1254

Qy       1048 -STNPVPTVATST-LASASEEDNKSGSHESAST-------------SLK--------PSMG 1085
                   || || | :    :    || | | :||              | |         |  |
Db       1255 VVRTTAPTTATPTPIIGDNNGSGKSKSGELSSTGSVTTNTATPDVPSTKVPSNPGAPGTG 1314

Qy       1086 ENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDT--------------REQPTTL 1131
                  |  ||| : |    |      ||  : ||:  | |                |
Db       1315 VPPPLAPSTETQTTNNVPGSPNIPATGT-TDIIRESTTVSHTVTGNGNTGVPMNPNPALT 1373

Qy       1132 STTSKTNSELVATTQATNE------NGGKSPSTDLTS--SLTTGTSASTSANSELVTSGS 1183
              ::|| |    .||  |||       | |   ||:: :  |  | |      ::   | |
Db       1374 TSTSLTG----ATNSATNPSHETGVNTGSRGSTNIVTPPSSATATVVIPGTDNGATTKGQ 1429

Qy       1184 VTGGAVAS------------------ASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSS 1224
               | |  ::                  :|        | : |: | :  |||||| |||
Db       1430 DTAGGNSNGFTATTNTQGGNNEPGNQPGTNTTGEPVGTTGTQSVESTSQPTTLSQQTTSS 1489

Qy       1225 SPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLF 1259
                ||   :|||:||||||:||| ||| |:| :|:|
Db       1490 LISTP--LASTFDGSGSIVQHSGWLYVLLTAISIF 1522


RESULT 3
Q9Y8F2
ID   Q9Y8F2      PRELIMINARY;      PRT;  1443 AA.
AC   Q9Y8F2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Agglutinin-like protein 6.
GN   ALS6.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=20321177; PubMed=10861907;
RA   Hoyer L.L., Hecht J.E.;
RT   "The ALS6 and ALS7 genes of Candida albicans.";
RL   Yeast 16:847-855(2000).
DR   EMBL; AF075293; AAD42033.1; -.
DR   InterPro; IPR001092; HLH_basic.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ   SEQUENCE   1443 AA;  151476 MW;  DCBD693F4C435809 CRC64;

  Query Match             43.0%;  Score 2791.5;  DB 3;  Length 1443;
  Best Local Similarity   44.2%;  Pred. No. 2.7e-109;
  Matches  642;  Conservative  202;  Mismatches  387;  Indels  223;  Gaps   35;

Qy          8 LFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTF 67
              || | :|| ||||:||| |||||::|  || : ||||||| ||||||||| |:|||||
Db          9 LFFYCTIAMAKTISGVFTSFNSLTYTNTGNYPYGGPGYPTWTAVLGWSLDGTLASPGDTF 68

Qy         68 TLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGT 127
              || |||||: |:|||||||:||||||| |::||:|||||:::| ||: | |:|:||||
Db         69 TLVMPCVFKFITTQTSVDLTANGVKYATCTFHAGEDFTTFSSMSCVVNNGLSSNIRAFGT 128

Qy        128 VTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYAS 187
              | |||:|||||||| :::|||||||||||||| |||  ||  | | |:    |: :| :
Db        129 VRLPISFNVGGTGSSVNIQDSKCFTAGTNTVTFTDGDHKISTTVNFPKTPQSSSSLVYFA 188

Qy        188 RVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSS 247
              ||:|||:|:::| || ||  || || :|||::  || |||| ||:||| ||| || ||||
Db        189 RVIPSLDKLSSLVVASQCTAGYASGVLGFSATKDDVTIDCSTIHVGITNGLNSWNMPVSS 248

Qy        248 ESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYI--SATDVNQYTLAYTNDYTCAGSRLQS 305
              |||||||||| |   | |:|||||||||||:|:  |||  | : | ||| | |   :  :
Db        249 ESFSYTKTCTPNSFIITYENVPAGYRPFIDSYVKKSATATNGFNLNYTNIYNCMDGKKGN 308

Qy        306 KPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQPIPTT 365
              |    || | |||||||| :| ||||||||||:|||||:|:||||||||:::|||||
Db        309 DPLIYFWTSYTNSDAGSNGAAVVVTTRTVTDSTTAITTLPFDPTVDKTKTIEVIEPIPTT 368

Qy        366 TITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSID 425
              |||||||||::||  |||| || ||||:|||||||||:|| |||: ||::| ||||||||
Db        369 TITTSYVGISTSLSTKTATIGGTATVVVDVPYHTTTTITSIWTGSATTSSTYTNPTDSID 428

Qy        426 TVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFA 485
              ||||||| |||||:||::|| |  || |||  | |||:|||:|| | ||||||:||:|||
Db        429 TVVVQVPSPNPTVTTTQFWSGSVPTTETVTTGPQGTDSVIIKEPHNPTVTTTEFWSESFA 488

Qy        486 TTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPT 545
              || |||  | |||||||:|| ||||||||:||:|||||  |||  | ||||||:||| |||
Db        489 TTETVTNNPEGTDSVIIKEPHNPTVTTTEFWSESFATTETVTNYPEGTDSVIVREPHNPT 548

Qy        546 VTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGT 605
              |||||:||:|:||| |||  | ||||||:||| | ||||||:||:|:||| |:|  | ||
Db        549 VTTTEFWSESFATTETVTNYPEGTDSVIVREPHNPTVTTTEFWSESFATTETITNYPEGT 608

Qy        606 DTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREP---------------- 649
              |:||:||| | ||||||:||:|| || |:|  | |||:::| :|
Db        609 DSVIVREPHNPTVTTTEFWSESFVTTETITTGPLGTDSIVIHDPLEESSSSTAIESSDSN 668

Qy        650 ---------------------------------PNPTVTTTEYWS------QSYATTTT 669
                                              |: ::  |   |       ||:::|:
Db        669 ISSSAQESSSSVEQSLTSADETSSIVELSSRSDIPSSSIGLTSSESSTVSSYDSYSSSTS 728

Qy        670 ITAPPGETDTVLIREPPNHTVT-----------TTEYWSQS----YATTTTVTAPPGETD 714
               ::     |:       : |::           || :|  |     :|: | ::
Db        729 ESSIASSYDSYSSSSIESSTLSSSDRCSSSISDTTSFWDSSSSDLESTSITWSSSIDAQS 788

Qy        715 TVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREP------------PNPTV 762
              : |::   |   |: |  | |   ::|       || ::  :              |: |:
Db        789 SHLVQSVSNSISTSQELSSSSSEESSTF-----ATDALVSSDASSILSSDTSSYYPSSTI 843
```

```
Qy        763 TTTEYW-----------SQSFATTT--------TVTAPPGGT-DTVIIYESMSSSKISTS 802
              :::: :           | || |:|        ::|: |  : |:     | |||  |:
Db        844 SSSDDFPHTIAGESDSLSISFITSTVEISSDSVSLTSDPASSFDSSSSLNSDSSSSPSSD 903

Qy        803 SNDI-------TSIIPSFS----------RPHYVNSTTSDLSTFESSSMNTPTSISSDGM 845
              :||       | ::|||           ||||||||   |  ||||: :|:  |
Db        904 QSDILTSSSFSTLVVPSFSLSSSSSLSLTYPHYVNSTTYHASESESSSVASPSMASESAN 963

Qy        846 LLSSTTLVTESETTTELICSDGKE---CSRLSSSSGIVTN-PDSNESSIVTSTVPTASTM 901
                  |  ::||  ||  | :|      | | :    |||  | |:  |  || | | |:
Db        964 --DDTYTLSESTDTTSSIGTDSSTVTFCRRDNGDGCIVTGMPSSSIDSEQTSDVTTTSSF 1021

Qy        902 SDSLSSTDG-ISATSSDNV--SKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSV 958
              | : |    | | : |:  |::  | :|: ||:   |  | :| | ||   |||  |
Db       1022 VASSTPTSAEQSITDNPNIDSSQTSASSSTKLSVSVSDTVVNSISLSETS--TLSSDDST 1079

Qy        959 SESESKVTFTSNGD----NQSGTHDSQSTSTEIEIVTTS---STKVLPPVVSSNTDLTSE 1011
              | |:: :: |:| |    |   :| | ::  |  | |    |:  |   :|| :|:|:|
Db       1080 S-SDTSISSTTNSDTGNVNAGSSHTSTASIKESSIQKTGVTLSSSYLSTKLSSTSDITTE 1138

Qy       1012 PTNTREQPTTLSTTSNSITED----ITTSQP-------TGDNG------DNTSSTNPVPT 1054
                |    | |:||  :  ||      || |       : ||       |  |:    |
Db       1139 LITTELITTELTTTELTTIEDNEPNTFTSTPSSHSEIFSSDNSVLSKQVDGESTVEIPPV 1198

Qy       1055 VATSTLASASEEDNKSGSHESASTSLKPSMGENS-------GLTTSTEIEATTTSPTEA- 1106
              |:|::| |    :      |||:   ::|||         : | | : :|::|   |
Db       1199 TDTTTVSSVSVHSTE------ASTA---TLGENSFSKVASAPVNTETSLRSTSSSSNHAT 1249

Qy       1107 --------------PSPAVSSGT--DVTTEPTDTREQPTTLSTTS-KTNSELVATTQAT 1148
                           |||  |:     :||       | : || :  | |:: | | :|
Db       1250 ESSGTVKSEASAEAIPSPPTSTDNRLSYSTEEAKGSTYPNSGSTNNLMTESQVAAPTDST 1309

Qy       1149 N---ENGGKSPSTDLTSSLTTGTSASTSANSELVTS----GSVTGGAVASASNDQS---- 1197
              :    |   : : |  ||      : | :  | : |          | :|: |  ::
Db       1310 SVLTANPVVTSTFDDKSSAAVNQPSKTKSIEESIGSLDSVNETNNGFIATLSQSEAPNSL 1369

Qy       1198 -HSTSVT-----------NSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQH 1245
               || |::           | :|  ||:  ||| ||| :|  | |  : :|| || |  :|
Db       1370 IHSESISTTMAKTTDASINGDSAASNSQPTTLIQQVATS--SYNQPLITTYAGSSSATKH 1427

Qy       1246 STWLYGLITLLSLF 1259
              :||   |::   |
Db       1428 PSWLLKFISVALFF 1441
```

```
RESULT 4
Q9Y743
ID   Q9Y743      PRELIMINARY;      PRT;  1270 AA.
AC   Q9Y743;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Agglutinin-like protein.
GN   ALS5.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=21064501; PubMed=11124701;
RA   Hoyer L.L., Hecht J.E.;
RT   "The ALS5 gene of Candida albicans and analysis of the Als5p N-
RT   terminal domain.";
RL   Yeast 18:49-60(2001).
DR   EMBL; AF068866; AAD32849.1; -.
SQ   SEQUENCE   1270 AA;  133337 MW;  D8E5FA5853F6D5C5 CRC64;
```

Query Match 41.2%; Score 2673.5; DB 3; Length 1270;
Best Local Similarity 45.9%; Pred. No. 2e-104;
Matches 629; Conservative 187; Mismatches 340; Indels 215; Gaps 39;

```
Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              |:|||||||||| |:|| |||:|:| :||||||| ||||||||||||||||||||||||
Db          1 MIQQFTLLFLYLSFATAKAITGIFNSIDSLTWSNAGNYAFKGPGYPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              ||||||| |||||||||:| || ||||||||||||||||||||||||||||||||||||
Db         61 ANPGDTFILNMPCVFKFTASQKSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              ||||||||||||||||||||||||||||||||||||||||| |||||||| | :|| | |||||||
Db        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGINTVTFNDGSKKLSIAVNFEKSTVDR 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              | ||  || |||||: ||:|||||||||||||||:| |||||||||:||||:||:||
Db        181 SGYLTTSRFMPSLNKIATLYVAPQCENGYTSGTMGFSTSYGDVAIDCSNVHIGISKGVND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||:||:|||||||||:|:| || | |||||||||||||||| :| ||| |:| |||||
Db        241 WNHPVTSESFSYTKSCSSFGISITYQNVPAGYRPFIDAYISPSDNNQYQLSYKNDYTCVD 300

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
                 |  ||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DYWQHAPFTLKWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
              |||||||||||||||||| ||||||||||:|||||||||||||:|||||| |||||||
Db        361 PIPTTTITTSYVGVTTSYSTKTAPIGETATLIVDVPYHTTTTVTSKWTGTITMTTTRTNP 420

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
              |||||||||||||||| :||::||:|| :|||:|  | |||:||::|| | ||||||:|
Db        421 TDSIDTVVVQVPLPNPTTTTTQFWSESFTSTTTITNKPEGTDSVIVKEPHNPTVTTTEFW 480

Qy        481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTT------------EYWSQSFATTTTVT 527
              |:|:||| |:|  | ||||::| :|   : :||            :  | |   ::::
Db        481 SESYATTETITTGPLGTDSIVIHDPLEESSSTTAIESSDSNISSSAQESSSSVEQSSSIV 540

Qy        528 APPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEY 587
                  :|  :  : |:             :::| :|:    | |    :  : :  :|
Db        541 GLSSSSDIPLSSDMPS------------SSSTGLTSSESSTVSSYDSDSSSSSELSTFS 587

Qy        588 WSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW--SQSFATTTTVTGPPSGTDTVI 645
              |:||::: :                    || :|  | |   :|::|   |
Db        588 SSESYSSSIS--------------------DTTNFWDSSSSDLESTSITWSSS------ 620

Qy        646 IREPPNPTVTTTEYWSQSYATTTTITAPPGE------TDTVLIREPPNHTVTTTEYWSQS 699
              |    : :| :    | | :|:   |:  ||      || :|:    :  :  :      |
Db        621 IDAQSSQSVQSV---SNSISTSQETTSSSGEESNTSVTD-ILVSSDASSILNSD---ISS 673

Qy        700 YATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPN 759
              |  ::|::       |:   || : :       | | |:|  ::     :| | :   |
Db        674 YYPSSTISLSDDFPHTI-AGEPDSRS-------SSSIASTVEIS-----SDLVSLTSDP- 719

Qy        760 PTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSIIPSFS----- 814
                |: : | |   : : ::|                | || ||:  | : ||||
Db        720 ----TSSFDSSSSLNSDSSSSP----------FSDESDISASSSFSTLVAPSFSLSSSS 764

Qy        815 -----RPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKE 869
                    ||||||||   |  ||||: :| |::|:     : || :|  || :| :|
Db        765 SLSLIYPHYVNSTTYHASESESSSVASP-SVASESANDDTHTL-SESTDTTSIIGTDSST 822

Qy        870 ---CSRLSSSSGIVT-----NPDSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNV-- 919
                | | :    |||      : || ::| ||:   |:| :: |  ::   | | : |:
Db        823 VTFCRRDNGDGCIVTGITSSSIDSEQTSDVTT---TSSFVASSTPTSAEQSITDNPNIDS 879

Qy        920 SKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGD----NQS 975
              |::  | :|::||:   |  | :  | ||   |||    | |:: :: |:| |    |
Db        880 SQTSASSSTKSSVSVSDTVVNSILLSETS--TLSSDDGTS-SDTSISSTTNSDTGNINAG 936

Qy        976 GTHDSQSTSTEIEIVTTS---STKVLPPVVSSNTDLTSEPTNTR------EQPTTLSTTS 1026
              :| | ::  |  |  |    |:  |   :|| :|:| |  |        :| | ::|
Db        937 SSHKSTASIKESSIQKTGVTLSSSYLSTKLSSTSDITIELITTELTTIEDNEPNTFTSTP 996
```

```
Qy       1027 NSITEDITTSQPTGDNG------DNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080
Db        997 SSHSEIFSS-----DNSVLSKQVDRESTIKTSPTTDVTTVSSLSVHSTE------ASTA- 1044

Qy       1081 KPSMGENS-------GLTTSTEIEATTTSPTEAPSPAVSSGT--------DVTTEPTDT- 1124
Db       1045 --TLGENSFSNVASTPLNTATSLRSTSSSSNHATE---SSGTVKSEASVEAIPSPPTSTD 1099

Qy       1125 --------REQPTTLSTTSKTN-----SELVATTQATN---ENGGKSPSTDLTSSLTTGT 1168
Db       1100 NRLSYSTEEAEGITYANSGSTNNLITESQVAAPTDSTSVLIENLVVTSTFDDNSSAAVDQ 1159

Qy       1169 SASTSANSELV----TSGSVTGGAVASASNDQSHSTSV----------------TNSNSI 1208
Db       1160 PSKTKSIEESIMNPDSTNETNNGFIATLSQAQVPSSSIHSELISTTTAKTTDASMNGDSA 1219

Qy       1209 VSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLF 1259
Db       1220 ASNSQPTTLIQQVATS--SYNQPLITTYAGSSSATKHPSWLLKFISVALFF 1268


RESULT 5
Q9HGK6
ID   Q9HGK6      PRELIMINARY;      PRT;  2297 AA.
AC   Q9HGK6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Agglutinin-like protein Als7p.
GN   ALS7.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=20321177; PubMed=10861907;
RA   Hoyer L.L., Hecht J.E.;
RT   "The ALS6 and ALS7 genes of Candida albicans.";
RL   Yeast 16:847-855(2000).
DR   EMBL; AF201684; AAF98068.1; -.
SQ   SEQUENCE   2297 AA;  244723 MW;  59B020C63027F651 CRC64;

  Query Match             39.7%;  Score 2576.5;  DB 3;  Length 2297;
  Best Local Similarity   41.3%;  Pred. No. 4.3e-100;
  Matches  595;  Conservative  220;  Mismatches  416;  Indels  211;  Gaps   24;

Qy          2 LQQFTLLFLYLSIAS--AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGT 59
Db          1 MKKLYLLLYLLASFTTVISKEVTGVFNQFNSLIWSYTYRARYEEISTLTANAQLEWALDGT 60

Qy         60 SANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALK 119
Db         61 IASPGDTFTLVMPCVYKFMTYETSVQLTANSIAYATCDFDAGEDTKSFSSLKCTVTDELT 120

Qy        120 SSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVD 179
Db        121 EDTSVFGSVILPIAFNVGGSGSKSTITDSKCFSSGYNTVTFFDGNNQLSTTANFLPRREL 180

Qy        180 PSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLN 239
Db        181 AFGLVVSQRLSMSLDTMTNFVMSTPCFMGYQSGKLGFTSNDDDFEIDCSSIHVGITNEIN 240

Qy        240 DWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCA 299
Db        241 DWSMPVSSVPFDHTIRCTSRALYIEFKTIPAGYRPFVDAIVQIPTTEPPFVKYTNEFACV 300

Qy        300 GSRLQSKPFTLRWT-GYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEI 358
Db        301 NGIYTSIPFTSFFSQPILYDEALAIGADLVRTTSTVIGSITRTTTLPFISRLQKTKTILV 360

Qy        359 LQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRT 418
Db        361 LEPIPTTTVTTSHHGFDTWYYTKKATIGDTATVFIDVPQHTATTLTTYWQESSTATTTYF 420

Qy        419 NPTDSIDTVVVQVPLP----------------------------------NPTVSTTE 442
Db        421 DDIDLVDTVIVKIPYPNPTIITTQFWSGKYLTTETHKEPPLGTDSVIIKEPHNPTVTTTE 480

Qy        443 YWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVII 502
Db        481 FWSESFATTETITNYPEGTDSVIVREPHNPTVTTTEFWSESFATTETITNGPEGTDSVIV 540

Qy        503 REPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTV 562
Db        541 REPHNPTVTTTEFWSESFATTETITNGPEGTDSVIVREPHNPTVTTTEFWSESFATTETI 600

Qy        563 TAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTE 622
Db        601 TNGPEGTDSVIVREPHNPTVTTTEFWSESFATTETITNGPEGTDSVIIKEPHNPTVTTTK 660

Qy        623 YWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLI 682
Db        661 FWSESFATTETITNYPEGTDSVIVREPHNPTVTTTEFWSESFATTETITNGPEGTDSVIV 720

Qy        683 REPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTV 742
Db        721 REPHNPTVTTTEFWSESFATTETITNYPEGTDSVIVREPHNPTVTTTEFWSESFATTETV 780

Qy        743 TAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIY----ESMSSSK 798
Db        781 TNYPEGTDSVIVREPHNPTVTTTEFWSESFVTTETITTGPLGTDSIVIHDPLEESSSSTA 840

Qy        799 ISTSSNDITSI----------------------------IPS-------------FSRP 816
Db        841 IESSDSNISSSAQESSSSVEQSFTSADETSSIVELSSRSDIPSSSIGLTSSESSTVSSYD 900

Qy        817 HYVNSTT-----SDLSTFESSSMNTPTSISSD--GMLLSSTT------------------ 851
Db        901 SYSSSTSESSIASSYDSYSSSSIESSTLSSSDRYSSSISDTTSFWDSSSSDLESTSITWS 960

Qy        852 ------------LVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAS 899
Db        961 SSIDAQSSHLVQSVSNSISTSQEISSSSSEESSTSATDALVSSDASSILSSDTSSYYPSS 1020

Qy        900 TMSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQL----SSI 955
Db       1021 TISPSDDFPHTI-AGESDSQSISFITSTVEISSDSVSLTSDP-ESSFDSSSRLNSDSSSS 1078

Qy        956 PS------VSESESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPP--------- 1000
Db       1079 PSTDQRDILTSSSFSTLIKSSGSRESSIGTILSEESSDSIPTTFSTRYWSPSGMSSRHYT 1138

Qy       1001 ---------VVSSNT--DLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSST 1049
Db       1139 NSTETSVSDVVSSSVAGDETSESSVSVISESSESVTSESVASESVASESVASESVASESV 1198

Qy       1050 NPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSP 1109
Db       1199 TAVSDI--SDLYTTSEEVSTSDSNSGMSSPI-PSSEQRS------SIPIMSSSDESSESR 1249

Qy       1110 AVSSGTDVTTEPTDTREQPTTLST----------------TSKTNSELVATTQATNENGG 1153
Db       1250 ESSSGTILSEENSDS--IPTTFSTRYWSPSGMSSRHYTNSTETSVSDVVSSSVAGDETSE 1307

Qy       1154 KSPS--TDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTNSNS---- 1207
Db       1308 SSVSVISESSESVTSESVASESVASESVASESVTAVSDISDLYTTSEVVSTSDSNSGMSS 1367

Qy       1208 ------------IVSNTPQ----------TTLSQQVTSSSPSTNTFIASTYDGSGSIIQH 1245
Db       1368 PIPSSEQRSSIPVMSSSDESSESRESSSGTILSEENSDSIPT--TFSTRYLSPSGMSSRH 1425

Qy       1246 ST 1247
```

```
Db       1426 YT 1427


RESULT 6
Q9URQ0
ID   Q9URQ0      PRELIMINARY;      PRT;   468 AA.
AC   Q9URQ0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   Agglutinin-like protein (Fragment).
GN   ALS2.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98440424; PubMed=9765564;
RA   Hoyer L.L., Payne T.L., Hecht J.E.;
RT   "Identification of Candida albicans ALS2 and ALS4 and localization of
RT   als proteins to the fungal cell surface.";
RL   J. Bacteriol. 180:5334-5343(1998).
DR   EMBL; AF024582; AAC64237.1; -.
FT   NON_TER     468    468
SQ   SEQUENCE   468 AA;  50143 MW;  2E38E36D7FEAD192 CRC64;

  Query Match             30.6%;  Score 1986.5;  DB 3;  Length 468;
  Best Local Similarity   80.0%;  Pred. No. 4.1e-76;
  Matches  375;  Conservative   33;  Mismatches   60;  Indels    1;  Gaps    1;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              || || || | :|:|:|| |||||:||:||||: | |||:||| ||||||||||||||
Db          1 MLLQFLLLSLCVSVATAKVITGVFNSFDSLTWTRAGNYAYKGPNRPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              |||||||||||||||: | ||||||||:||||||||||||||||||:| |||:: | |
Db         61 ANPGDTFTLNMPCVFKFITDQTSVDLTAEGVKYATCQFYSGEEFTTFSSLKCTVSNTLTS 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              |||| |||||||:|||||||||| ||| |:|| |||||||||||| |||||:|||:  |
Db        121 SIKALGTVTLPISFNVGGTGSSVDLESSQCFKAGTNTVTFNDGDKKISIDVDFEKTNEDA 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              | :  |||::||:|||:  :||||| |||||| |||    || |||||:|:|||||||
Db        181 SGFFIASRLIPSINKVSITYVAPQCANGYTSGAMGFIVLTGDTTIDCSNVHVGITKGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||:||||:| || |||:| || | |:||||||||| | | | :  |:  | ||||| | |
Db        241 WNFPVSSDSLSYNKTCSSTGISITYENVPAGYRPFFDVYTSVSGQNR-QLRYTNDYACVG 299

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
              | |||||| ||  || ||:| ||| ||||||||||||||||||||||||||||||||||
Db        300 SSLQSKPFNLRLRGYNNSEANSNGFVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 359

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
              |||||||||||||||| ||||||||||||||||||||||||||||||| ||||||||
Db        360 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITATTTRTNP 419

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREP 469
              |||||||||||| |||||||||||||:|||||||||||||||:||||||
Db        420 TDSIDTVVVQVPSPNPTVSTTEYWSQSYATTTTVTAPPGGTDSVIIREP 468


RESULT 7
Q9URP8
ID   Q9URP8      PRELIMINARY;      PRT;   469 AA.
AC   Q9URP8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   Agglutinin-like protein (Fragment).
GN   ALS4.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=98440424; PubMed=9765564;
RA   Hoyer L.L., Payne T.L., Hecht J.E.;
RT   "Identification of Candida albicans ALS2 and ALS4 and localization of
RT   als proteins to the fungal cell surface.";
RL   J. Bacteriol. 180:5334-5343(1998).
DR   EMBL; AF024586; AAC64241.1; -.
FT   NON_TER     469    469
SQ   SEQUENCE   469 AA;  49597 MW;  88BC96D79142C8DB CRC64;

  Query Match             28.1%;  Score 1823;  DB 3;  Length 469;
  Best Local Similarity   72.1%;  Pred. No. 2.8e-69;
  Matches  338;  Conservative   49;  Mismatches   82;  Indels    0;  Gaps    0;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              || || || | :|:|:|| |||||:||:||||:|||:| ::||  ||| ||:|||||| :
Db          1 MLLQFLLLSLCVSVATAKVITGVFNSFNSLTWANAASYPYRGPATPTWTAVIGWSLDGAT 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              |: ||||||:||||||: | |||:|| |||  ||||   | |||||||:::|||   : :
Db         61 ASAGDTFTLDMPCVFKFITDQTSIDLVADGRTYATCNLNSAEEFTTFSSVSCTVTTTMTA 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              || |||||| :|:|||:|| || :|:||||| |||||||||  ||  |:||||||
Db        121 DTKAMGTVTLPFSFSVGGSGSDVDLANSQCFTAGINTVTFNDGDTSISTTVDFEKSTVAS 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              |  :  ||::|||::  :||:  :| ||||||||||::    |||| :|:||: ||||
Db        181 SDRILLSRILPSLSQAVSLFLPQECANGYTSGTMGFSTAGTGATIDCSTVHVGISNGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||||:|||||||||||  : : ||||||||||:|||:||| |: ||: ||| | | |
Db        241 WNYPISSESFSYTKTCTSTSVLVTYQNVPAGYRPFVDAYVSATRVSSYTMQYTNIYACVG 300

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
              :      ||  | || || |||||| || |||||||||||||||||   ||||||||||
Db        301 AASVDDSFTHTWRGYSNSQAGSNGITIVVTTRTVTDSTTAVTTLPFNSDTDKTKTIEILQ 360

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREP 469
              |||||||||||| |||||:|||||||:|||||||||||||:||||||
Db        421 TDSIDTVVVQVPSPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREP 469


RESULT 8
Q9C471
ID   Q9C471      PRELIMINARY;      PRT;   468 AA.
AC   Q9C471;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Agglutinin-like protein (Fragment).
GN   ALS9.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RA   Hoyer L.L., Hecht J.E., Mirus K.A.;
RT   "The ALS9 gene of Candida albicans.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
```

```
DR   EMBL; AF229989; AAK00764.1; -.
FT   NON_TER     468    468
SQ   SEQUENCE   468 AA;  50127 MW;  B291D3EB15FB96DE CRC64;

  Query Match             27.4%;  Score 1780.5;  DB 3;  Length 468;
  Best Local Similarity   70.1%;  Pred. No. 1.7e-67;
  Matches  329;  Conservative   62;  Mismatches   77;  Indels    1;  Gaps    1;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              || || |||: |::::|||||||||:||:||||: :  ||:|||  ||||||||||: |:
Db          1 MLPQFILLFISLTVSTAKTITGVFNSFDSLTWTRSVEYAYKGPETPTWNAVLGWSLNSTT 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              |:|||||| :|||||: |:|||||||||| |||| | :|||||||:|:|||||    |
Db         61 ADPGDTFTLILPCVFKFITTQTSVDLTADGVSYATCDFNAGEEFTTFSSLSCTVNSVSVS 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
                :  ||| ||| |||||||| || |||||||| |||| ||| ||  |:|: | | |
Db        121 YARVSGTVKLPITFNVGGTGSSVDLADSKCFTAGKNTVTFMDGDTKISTTVDFDASPVSP 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              | |: :||::|||||:::||| |||||||||| ||| :|||   |||||::|||:|||||
Db        181 SGYITSSRIIPSLNKLSSLFVVPQCENGYTSGIMGFVASNG-ATIDCSNVNIGISKGLND 239

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||:|||||||||||||  | :::|||||||||:|||||| ::::||| | |:|||
Db        240 WNFPVSSESFSYTKTCTSTSITVEFQNVPAGYRPFVDAYISAENIDKYTLTYANEYTCEN 299

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
                  |||| | |||||:| |:| ||| |||||||||||||||||||||:||||||
Db        300 GNTVVDPFTLTWWGYKNSEADSDGDVIVVTTRTVTDSTTAVTTLPFNPSVDKTETIEILQ 359

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
              |||||||||||:|::||| | |  || ||||||| ||| | |||: |||:|||||| |||
Db        360 PIPTTTITTSYIGISTSYETLTGTIGGTATVIVDTPYHITATVTNFWTGSITTTTTYTNP 419

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREP 469
              | |||||:||:| |:|| : ||:||:|||:|||:| || ||::|||:||
Db        420 TGSIDTVIVQIPSPDPTTTITEFWSESFASTTTITNPPDGTNSVIIKEP 468


RESULT 9
Q9HF70
ID   Q9HF70      PRELIMINARY;      PRT;   336 AA.
AC   Q9HF70;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Agglutinin-like protein Alsd2p (Fragment).
GN   ALSD2.
OS   Candida dubliniensis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=42374;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD36;
RA   Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;
RT   "Evidence suggesting the presence of an ALS gene family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF202529; AAG35623.2; -.
FT   NON_TER       1      1
FT   NON_TER     336    336
SQ   SEQUENCE   336 AA;  35593 MW;  E513E6EA9E8E9EC7 CRC64;

  Query Match             22.6%;  Score 1471;  DB 3;  Length 336;
  Best Local Similarity   81.0%;  Pred. No. 1.1e-54;
  Matches  272;  Conservative   31;  Mismatches   33;  Indels    0;  Gaps    0;

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              |:|||||| |||||: |:|||||||:|||||| ||||||||||:|||||| || |
Db          1 ASPGDTFTLIMPCVFKFITTQTSVDLTANGVKYATCSFYSGEEFTTFSSLTCTVNSALTS 60

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              |:||||||||||:|||||||| ||||||||||||||||| |||  :|  |:|||||
Db         61 SVKAFGTVTLPISFNVGGTGSSVDLEDSKCFTAGTNTVTFTDGDNKVSTTVDFEKSTVDS 120

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              : || :||:|||||||:|||||| ||||||:||||||  |: ||||:|:||| |:||
Db        121 TGYLTSSRLMPSLNKVTSLFVAPQCARGYTSGTIGFSSSNGGVSFDCSNVHVGITNGVND 180

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||:|||||||||||:| || | |:|||||||||||:|:||| |||:|||||| |
Db        181 WNFPVSSESFSYTKTCSSTGITITYRNVPAGYRPFIDAYITASDVNSYTLSYTNDYTCVG 240

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
                :| |||||||:|||| :|||||||||||||||||||||||||  ||||||||||
Db        241 GSVQHKPFTLRWSGYKNGEAGSNGIVIVATTRTVTDSTTAVTTLPFNSDVDKTKTIEILQ 300

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVP 396
              |||||||| |||||||| |||||||:|||: :|:|
Db        301 PIPTTTITTLYVGVTTSYSTKTAPIGDTATLFIDIP 336


RESULT 10
Q9HF72
ID   Q9HF72      PRELIMINARY;      PRT;   338 AA.
AC   Q9HF72;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Agglutinin-like protein AlsD1p (Fragment).
GN   ALSD1.
OS   Candida dubliniensis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=42374;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD36;
RA   Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;
RT   "Evidence suggesting the presence of an ALS gene family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF201685; AAG35603.1; -.
FT   NON_TER       1      1
FT   NON_TER     338    338
SQ   SEQUENCE   338 AA;  35908 MW;  C0A7F2F94609E172 CRC64;

  Query Match             18.3%;  Score 1191;  DB 3;  Length 338;
  Best Local Similarity   64.8%;  Pred. No. 5.6e-43;
  Matches  219;  Conservative   51;  Mismatches   66;  Indels    2;  Gaps    1;

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              || |||||| ||||||: |:|||||||:|||||| |::||:|| ||:::| ||: | |
Db          1 ANAGDTFTLIMPCVFKFITTQTSVDLTANGVKYATCTFHAGEDFTAFSSMSCVVNNGLTS 60

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              :||||||| :||:|||||||||| :|:|||||||||:||| |||  ||| |:| |:
Db         61 NIKAFGTVRIPISFNVGGTGSSVNLQDSKCFTAGTNSVTFTDGDHKISIPVDFPKTPESS 120

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              |  :  |||:|:|:|:::| || ||  || || :|||::  || |:|||:|:||| |||
Db        121 SGLIKYSRVIPTLDKLSSLAVASQCTAGYKSGVLGFSATKNDVTIECSNVHVGITNGLNS 180

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQ--YTLAYTNDYTC 298
              || ||||:|||||||||:    | |:||||||||| |:  |      : | ||| | |
Db        181 WNMPVSSDSFSYTKTCTSSSFIITYENVPAGYRPFIDTYVKKTSTTSTGFNLNYTNSYVC 240

Qy        299 AGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEI 358
                :  : |    || | ||||||:| |:: ||:||||||||:|||||:|:||||||||:
Db        241 TDGKKGNDPLIYFWTSYTNSDAGSDGAVVIVTTKTVTDSTTAITTLPFDPTVDKTKTIEV 300

Qy        359 LQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVP 396
              |:|||||||||||||:||: |||| ||||||:|:|||
```

```
Db        301 LEPIPTTTITTSYVGVSTSFFTKTATIGETATLIIDVP 338


RESULT 11
Q9HF69
ID   Q9HF69      PRELIMINARY;      PRT;   331 AA.
AC   Q9HF69;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Agglutinin-like protein Alsd3p (Fragment).
GN   ALSD3.
OS   Candida dubliniensis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=42374;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CD36;
RA   Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;
RT   "Evidence suggesting the presence of an ALS gene family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF202530; AAG35624.2; -.
FT   NON_TER       1      1
FT   NON_TER     331    331
SQ   SEQUENCE   331 AA;  35297 MW;  469AB72F9CE029BF CRC64;

  Query Match             18.1%;  Score 1176;  DB 3;  Length 331;
  Best Local Similarity   64.3%;  Pred. No. 2.3e-42;
  Matches  214;  Conservative   44;  Mismatches   73;  Indels    2;  Gaps    1;

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              || ||||||||||||: | |||||| |||  |||| ||||||||||:| |||::|| |
Db          1 ANAGDTFTLNMPCVFKFITDQTSVDLVADGRTYATCDLYSGEEFTTFSSLKCTVSNALNS 60

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              || |||||::||:||:||  |:  |:||  ||||||||||    |    |::| |:
Db         61 QTKALGTVTLPLSFNIGGSGSDVDITSSQCFKEGTNTVTFNDGDTTFSTTANFQRSDVNA 120

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              :  :  ||::||| |  |:|: |:| :||:|||||||::  |  |||| :| ||: ||||
Db        121 NDRILLSRILPSLAKSVTIFIPPRCASGYSSGTMGFSTAGTDAIIDCSTVHAGISNGLND 180

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||||: |:||||| :|:| |: : ||||||||||:|||||  :  ||: ||| ||| |
Db        181 WNYPIESKSFSYTTSCSSKGVSVTYQNVPAGYRPFVDAYISA--LTSYTMQYTNQYTCVG 238

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
              :|     |:  | || |::||| || || || ||||||||||||||  ||:|||| :||
Db        239 ARPVDASFSYNWLGYDNAEAGSRGITIVVTTSTVTDSTTAVTTLPFNSDVDRTKTIAVLQ 298

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIV 393
              |||||||||||||||||| |:| |||||||: :
Db        299 PIPTTTITTSYVGVTTSYRTQTVPIGETATLFI 331


RESULT 12
Q9Y8F1
ID   Q9Y8F1      PRELIMINARY;      PRT;   433 AA.
AC   Q9Y8F1;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Agglutinin-like protein 7 (Fragment).
GN   ALS7.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1161;
RX   MEDLINE=20321177; PubMed=10861907;
RA   Hoyer L.L., Hecht J.E.;
RT   "The ALS6 and ALS7 genes of Candida albicans.";
RL   Yeast 16:847-855(2000).
DR   EMBL; AF075294; AAD42034.1; -.
FT   NON_TER     433    433
SQ   SEQUENCE   433 AA;  48037 MW;  1F30CFD99C2EC445 CRC64;

  Query Match             15.4%;  Score 999.5;  DB 3;  Length 433;
  Best Local Similarity   46.0%;  Pred. No. 7.3e-35;
  Matches  199;  Conservative   76;  Mismatches  155;  Indels    3;  Gaps    2;

Qy          2 LQQFTLLFLYLSIAS--AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGT 59
              :::  ||:|  |  :  :| :|||||: |||| ||      ::     | || | |:||||
Db          1 MKKLYLLYLLASFTTVISKEVTGVFNQFNSLIWSYTYRARYEEISTLTANAQLEWALDGT 60

Qy         60 SANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALK 119
               |:|||||| ||||:|: | :||| |||: : |||| | :||:  :||:| ||| | |
Db         61 IASPGDTFTLVMPCVYKFMTYETSVQLTANSIAYATCDFDAGEDTKSFSSLKCTVTDELT 120

Qy        120 SSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVD 179
                 ||:| |||||||||:|| : : |||||::| ||||| ||:  :|     |
Db        121 EDTSVFGSVILPIAFNVGGSGSKSTITDSKCFSSGYNTVTFFDGNNQLSTTANFLPRREL 180

Qy        180 PSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLN 239
                  : |:  ||: :|   ::  |  || || :|||:|:: |  ||||:||:|||  :|
Db        181 AFGLVVSQRLSMSLDTMTNFVMSTPCFMGYQSGKLGFTSNDDDFEIDCSSIHVGITNEIN 240

Qy        240 DWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCA 299
              ||: ||||  | :|  |||  : |::: :|||||||:|| :       : : |||:: |
Db        241 DWSMPVSSVPFDHTIRCTSRALYIEFKTIPAGYRPFVDAIVQIPTTEPFFVKYTNEFACV 300

Qy        300 GSRLQSKPFTLRWT-GYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEI 358
                  | |||  ::       :| : |  :| || ||  | |  |||||   : ||||| :
Db        301 NGIYTSIPFTSFFSQPILYDEALAIGADLVRTTSTVIGSITRTTTLPFISRLQKTKTILV 360

Qy        359 LQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRT 418
              |:|||||:|||: |  | | || | ||:|||| :||| || ||:|: |  : | |||
Db        361 LEPIPTTTVTTSHHGFDTWYYTKKATIGDTATVFIDVPQHTATTLTTYWQESSTATTTYF 420

Qy        419 NPTDSIDTVVVQV 431
              :  | :|||:|::
Db        421 DDIDLVDTVIVKI 433


RESULT 13
Q9HF71
ID   Q9HF71      PRELIMINARY;      PRT;   353 AA.
AC   Q9HF71;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   Agglutinin-like protein 1 (Fragment).
GN   ALST1.
OS   Candida tropicalis (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5482;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13803;
RA   Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;
RT   "Evidence suggesting the presence of an ALS gene family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF201686; AAG35604.1; -.
FT   NON_TER       1      1
FT   NON_TER     353    353
SQ   SEQUENCE   353 AA;  37272 MW;  4268154D5615DE08 CRC64;

  Query Match             15.1%;  Score 981;  DB 3;  Length 353;
  Best Local Similarity   54.9%;  Pred. No. 3.5e-34;
  Matches  185;  Conservative   50;  Mismatches  100;  Indels    2;  Gaps    2;
```

```
Qy          60 SANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALK 119
               :|| |||||| |||||||:|||:||:|||     |||| | :|| |||||:|:|||  ::
Db          19 AANAGDTFTLIMPCVFKFTTSETSIDLTVGSKSYATCNFNAGEHFTTFSSLSCTVTQSVP 78

Qy         120 SSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVD 179
                :  |:||:|:|:|||||||:|   || |:|||| | |||||:|||| |    || :
Db          79 DNTNAYGTITVPLAFNVGGSGRDVDLTDAKCFTTGDNTVTFSDGDKSFSTTANFEGAGTL 138

Qy         180 PSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLN 239
                 |  :|||:||| |   | ||| | ||| |||:|||:    :|||:||  |||  ||
Db         139 NDDY-ESSRLIPSLGKTDALLVAPLCSNGYKSGTIGFSSTTKGFSIDCNNIQAGITSQLN 197

Qy         240 DWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCA 299
                | :|  |:|||||  ||:    | :  :| | |||||||| |   : | : ||  | |:
Db         198 AWGFPTDSQSFSYTTQCTTTSYSITFSTIPKGLRPFIDAYIKA-PTSTYPMTYTYKYVCS 256

Qy         300 GSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEIL 359
               : :    | |:|| |||| | |: ||  | | | ||| |||||: : ||||||:::
Db         257 DGKSYNGNTKLNWSGYVNSDADSEGMEIVVATTTGTGSTTGVTTLPFDKTKDKTKTIQVI 316

Qy         360 QPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVP 396
               :||||||:||||:|||||: | || ||||||:::|:|
Db         317 EPIPTTTVTTSYLGVTTSFSTITATIGETATLVIDMP 353


RESULT 14
Q96WU8
ID   Q96WU8      PRELIMINARY;      PRT;  1195 AA.
AC   Q96WU8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical 119.8 kDa protein (Fragment).
GN   SPCPB16A4.07C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972H-;
RA   Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL591677; CAC39326.1; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW   Hypothetical protein.
FT   NON_TER    1195   1195
SQ   SEQUENCE   1195 AA;  119807 MW;  5D8892EAA086E835 CRC64;

  Query Match             12.8%;  Score 831.5;  DB 3;  Length 1195;
  Best Local Similarity   28.7%;  Pred. No. 2.2e-27;
  Matches  319;  Conservative  154;  Mismatches  446;  Indels  193;  Gaps   48;

Qy         251 SYTKTCT---SNGIQIKYQNVPA----GYRPFIDAYISATDVNQYTLAYTNDYTCAGSRL 303
               | | | |   |||  |    ||     ||      |  :| :   |: :: : |   :
Db           1 SVTTTLTGAPSNGTVIDTVEVPTTTNYGYTTVTTGYTGSTTLTT-TVPHSGNETGPTTVY 59

Qy         304 QSKPF-----TLRWTGYKNS------DAGSNGIVI------------VATTRTVTDSTT 339
                 |:     |    ||  |        | ||| ||                |  | |||
Db          60 VETPYPTTVTTTTTVGYPGSVTTTLTGAPSNGTVIDTVEIPTTTNYGYTTITTGYTGSTT 119

Qy         340 AVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVI--VDVPY 397
                ||:| : :     |: :  | |||  ||: ||   |  |         |||  |:||
Db         120 LTTTVPHSGNETGPTTVYVETPYPTTVTTTTTVGYPGSVTTTLTGAPSNGTVIDTVEVPT 179

Qy         398 HTT---TTVTSEWTGTITTTTT------RTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSF 448
                |    ||:|: :||: | |||       | ||    || |: | |    :||
Db         180 TTNYGYTTITTGYTGSTTLTTTVPHSGNETGPT----TVYVETPYPTTVTTTTTVGYPGS 235

Qy         449 ATTTTVTAPPGGT--DTVIIREPPN--HTVTTTEYWSQSFATTTTVTAPPGGTD----SV 500
                |||   ||  ||  ||| :    |  :|  || |   : :|| | | |  | :    :|
Db         236 VTTTLTGAPSNGTVIDTVEVPTTTNYGYTTVTTGY---TGSTTLTTTVPHSGNETGPTTV 292

Qy         501 IIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEY--------W 552
                :  |   ||||||      | : |||:|    |::|::       || |:  |       |
Db         293 YVETPYPTTVTTTTTVGYSGSVTTTLTG--SGSNSIVTETVDVPTTTSVNYGYTTITTGW 350

Qy         553 SQSYATTTTVT--APPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVII 610
                : |   |: ||       |  :| |  |     |||     | : |||:|   |  |  :
Db         351 TGSTTLTSIVTHSGSETGPTTVYIETPSVSATTTTTTIGYSGSLTTTLTGSSGPVVTNTV 410

Qy         611 REPPNH-------------TVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPT--- 653
                 |  :                |  || |     :| | :    :|| ||:|: |   |
Db         411 EIPYGNSSYIIPTTIVTGTVTTVTTGYTGTETSTVTVIPTGTTGTTTVVIQTPTTVTATE 470

Qy         654 ---VTTTEYWSQSYATTTTITAPPGE---TDTVLIREPPNHT------VTTTEYWSQSYA 701
                   || |  :: :  :| |:| | |    | ||:|: |   |      || |  :: :
Db         471 TDIVTVTTGYTGTETSTVTVT-PTGTSTGTTTVVIQTPTTVTATETDIVTVTTGYTGTET 529

Qy         702 TTTTVTAPPGE---TDTVLIREPPNHT------VTTTEYWSQSYATTTTVTAPPG---GT 749
               :| ||| | |    | ||:|: |   |      || |  :: :  :| ||| | |   ||
Db         530 STVTVT-PTGTSTGTTTVVIQTPTTVTATETDIVTVTTGYTGTETSTVTVT-PTGTSTGT 587

Qy         750 DTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSKISTSSNDITSI 809
                ||:|: |   | | |:          |||    ||:|    ::: :   ||:   | :
Db         588 TTVVIQTPTTVTATETD--------IVTVTTGYTGTET----STVTVTPTGTSTGTTTVV 635

Qy         810 IPSFSRPHYVNSTTSDLSTFESSSMNTPTS---ISSDGMLLSSTTLVTESETT-----TE 861
                | :   |  | :| :|: |  :    | ||   ::  |    :||:| :: ||      |:
Db         636 IQT---PTTVTATETDIVTVTTGYTGTETSTVTVTPTGTSTGTTTVVIQTPTTVTATETD 692

Qy         862 LICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAS---TMSDSLSSTDGISATSSDN 918
                ::           : |     |:| || | ||  :   | :  : :   ::|| :|
Db         693 IVT-------------VTTGYTGTETSTVTVT-PTGTSTGTTTVVIQTPTTVTATETDI 737

Qy         919 VSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTH 978
               |: :     ||||  |:     |  ::  : | : : |: : ||  |  |:      :||
Db         738 VTVTTGYTGTETSTVTV----TPTGTATGTTTVVINTPTTTGSE--VLPTTGATGTAGTE 791

Qy         979 DSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLS-TTSNSITEDITTSQ 1037
                  :|:||::  || :|        : | :|   | |  | || : | : : ||  ||:
Db         792 TQLTTATEVQ-PTTGAT----GTAGTETQVT---TGTETQATTATETQATTATEVQTTTG 843

Qy        1038 PTGDNGDNTSSTNPV---PTV-ATST------LASASEEDNKSGSHESASTSLKPSMGEN 1087
                ||  |  | :|      ||  || |       : :|:|    :|:  :| |  : : |
Db         844 ATGTAGTETQATTATEVQPTTGATGTAGTETQVTTATEVQPTTGATGTAGTETQVTTGTE 903

Qy        1088 SGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLST-TSKTNSELVATTQ 1146
                :  ||:|| :||| :  :  : | :     |  | |  |||| :| |: | :::   |:
Db         904 TQATTATETQATTATEVQTTTGATGTAGTETQATTATEVQPTTGATGTAGTETQVTTATE 963

Qy        1147 ---ATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVT 1203
                   |   |      | :|:   |  : :|   :   |    | || :| :   :    | ||
Db         964 VQPTTGATGTAGTETQVTTGTETQATTATETQATTATEVQTTTGATGTAGTE----TQVT 1019

Qy        1204 NSNSIVSNTPQTTLSQQVTSSSPSTNTFIAST 1235
                :  :     | | :   ||||    | ::||
Db        1020 TATEV-----QPTTAVTETSSSGYYTTIVSST 1046


RESULT 15
Q9HG16
ID   Q9HG16      PRELIMINARY;      PRT;   226 AA.
AC   Q9HG16;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   Agglutinin-like protein (Fragment).
GN   ALS5.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1177;
RA   Hoyer L.L., Kapteyn J.C., Hecht J.E., En J., Klis F.M.;
RT   "Evidence Suggesting the Presence of an ALS Gene Family in Candida
RT   dubliniensis and Candida tropicalis.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF189016; AAG17111.1; -.
FT   NON_TER       1      1
FT   NON_TER     226    226
SQ   SEQUENCE   226 AA;  24802 MW;  2AA36A42E44CB460 CRC64;

  Query Match             12.7%;  Score 824;  DB 3;  Length 226;
  Best Local Similarity   70.1%;  Pred. No. 8.3e-28;
  Matches  155;  Conservative   28;  Mismatches   38;  Indels    0;  Gaps    0;

Qy        429 VQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTT 488
              |||| ||||||:||::|| |  || ||| | |||:|||:|| | ||||||:||:|||||
Db          1 VQVPSPNPTVTTTQFWSGSVPTTETVTTGPQGTDSVIIKEPHNPTVTTTEFWSESFATTE 60

Qy        489 TVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTT 548
              |||  | ||||||||:|| |||||||||:||:||||| |||  | ||||||:||| ||||||
Db         61 TVTNNPEGTDSVIIKEPHNPTVTTTEFWSESFATTETVTNYPEGTDSVIVREPHNPTVTT 120

Qy        549 TEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTV 608
              ||:||:|:||| |||  | ||||||:||| | ||||||:||:|:||| |:|  | |||:|
Db        121 TEFWSESFATTETVTNYPEGTDSVIVREPHNPTVTTTEFWSESFATTETITNYPEGTDSV 180

Qy        609 IIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREP 649
              |:||| | ||||||:||:|| || |:|  | |||:::| :|
Db        181 IVREPHNPTVTTTEFWSESFVTTETITTGPLGTDSIVIHDP 221


Search completed: June 11, 2003, 17:15:00
Job time : 61 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2003, 17:11:32 ; Search time 29 Seconds
(without alignments)
4176.875 Million cell updates/sec

Title: US-09-715-876-8
Perfect score: 6495
Sequence: 1 MLQQFTLLFLYLSIASAKTI.....SIIQHSTWLYGLITLLSLFI 1260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6495	100.0	1260	2	S60896	agglutinin-like pr
2	3194	49.2	1419	2	T30531	agglutinin-like ad
3	939	14.5	1367	2	S51959	hypothetical prote
4	939	14.5	1537	2	S53465	flocculation prote
5	711	10.9	1075	2	S48992	flocculation prote
6	675.5	10.4	1367	1	S48478	glucan 1,4-alpha-g
7	649.5	10.0	1275	2	T33369	hypothetical prote
8	646.5	10.0	2232	2	T34434	hypothetical prote
9	637.5	9.8	2271	2	F90073	hypothetical prote
10	632.5	9.7	3570	2	T45025	mucin MUC5B, trach
11	599	9.2	3020	2	A43932	mucin 2 precursor,
12	578	8.9	1283	2	T39174	hypothetical Serin
13	577	8.9	1161	2	S57180	probable membrane
14	553.5	8.5	1459	2	T32271	hypothetical prote
15	547.5	8.4	1609	2	S25345	probable membrane
16	526	8.1	1229	2	T25697	hypothetical prote
17	526	8.1	1630	2	A53577	ascites sialoglyco
18	515	7.9	4776	2	E95206	cell wall surface
19	511.5	7.9	3507	2	T34513	hypothetical prote
20	503.5	7.8	1777	2	T34369	hypothetical prote
21	501	7.7	1169	2	S38181	flocculation prote
22	492.5	7.6	1104	2	S59310	probable membrane
23	487	7.5	1832	2	T31113	mucin-like glycopr
24	482.5	7.4	1829	2	T24583	hypothetical prote
25	464.5	7.2	670	2	F36791	hypothetical prote
26	460	7.1	770	2	T22808	hypothetical prote
27	451.5	7.0	786	2	T16509	hypothetical prote
28	445	6.9	1802	2	S69703	HKR1 protein precu
29	443.5	6.8	1032	2	T34433	hypothetical prote
30	443.5	6.8	1251	2	T21389	hypothetical prote
31	435.5	6.7	2468	2	A83412	hypothetical prote
32	430.5	6.6	866	2	T45462	membrane glycoprot
33	428	6.6	867	2	T45463	membrane glycoprot
34	424	6.5	528	2	I47141	gastric mucin (clo
35	423.5	6.5	520	2	S62521	hypothetical prote
36	419	6.5	973	2	T40778	hypothetical 129.5
37	418.5	6.4	725	2	A41258	a-agglutinin core
38	415.5	6.4	1306	2	S25370	MSB2 protein - yea
39	412.5	6.4	3029	2	S76109	hypothetical prote
40	412	6.3	851	2	T22696	hypothetical prote
41	409.5	6.3	1041	2	S55862	probable membrane
42	405	6.2	833	2	E90577	lipoprotein vsaC [
43	404.5	6.2	2035	2	A40718	host cell factor C
44	401.5	6.2	648	2	PC4395	mucin 3 - human (f
45	400.5	6.2	948	2	T11678	hypothetical prote

ALIGNMENTS

RESULT 1
S60896
agglutinin-like protein - yeast (Candida albicans)
C;Species: Candida albicans
C;Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 17-Mar-2000
C;Accession: S60896
R;Hoyer, L.L.; Scherer, S.; Shatzman, A.R.; Livi, G.P.
Mol. Microbiol. 15, 39-54, 1995
A;Title: Candida albicans ALS1: domains related to a Saccharomyces cerevisiae sexual
A;Reference number: S60896; MUID:95272392; PMID:7752895
A;Accession: S60896
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1260 <HOY>
A;Cross-references: EMBL:L25902; NID:g704426; PIDN:AAC41649.1; PID:g704427
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosid

Query Match 100.0%; Score 6495; DB 2; Length 1260;
Best Local Similarity 100.0%; Pred. No. 1.7e-273;
Matches 1260; Conservative .0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60

Qy         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120

Qy        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180

Qy        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240

Qy        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300

Qy        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

Qy        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
```

```
Qy       481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
Db       481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540

Qy       541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
Db       541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600

Qy       601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660
Db       601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYW 660

Qy       661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720
Db       661 SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIRE 720

Qy       721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780
Db       721 PPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTA 780

Qy       781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840
Db       781 PPGGTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSI 840

Qy       841 SSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900
Db       841 SSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTAST 900

Qy       901 MSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE 960
Db       901 MSDSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE 960

Qy       961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT 1020
Db       961 SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPT 1020

Qy      1021 TLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080
Db      1021 TLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSL 1080

Qy      1081 KPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE 1140
Db      1081 KPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSE 1140

Qy      1141 LVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHST 1200
Db      1141 LVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHST 1200

Qy      1201 SVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db      1201 SVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
```

RESULT 2
T30531
agglutinin-like adhesin - yeast (Candida albicans)
C;Species: Candida albicans
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Mar-2000
C;Accession: T30531
R;Gaur, N.K.; Klotz, S.A.
Infect. Immun. 65, 5289-5294, 1997
A;Title: Expression, cloning, and characterization of a Candida albicans gene, ALA1, tha
A;Reference number: Z20847; MUID:98053977; PMID:9393828
A;Accession: T30531
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1419 <GAU>
A;Cross-references: EMBL:AF025429; NID:g2522218; PID:g2522219; PIDN:AAB88883.1
C;Genetics:
A;Gene: ALA1
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

```
  Query Match             49.2%;  Score 3194;  DB 2;  Length 1419;
  Best Local Similarity   49.5%;  Pred. No. 1e-130;
  Matches  708;  Conservative  182;  Mismatches  355;  Indels  186;  Gaps   29;

Qy         1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
Db         1 MIQQFTLLFLYLSFATAKAITGIFNSIDSLTWSNAGNYAFKGPGYPTWNAVLGWSLDGTS 60

Qy        61 ANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKS 120
Db        61 ANPGDTFILNMPCVFKFTASQKSVDLTADGVKYATCQFYSGEEFTTFSSLKCTVNNNLRS 120

Qy       121 SIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180
Db       121 SIKALGTVTLPIAFNVGGTGSSVDLEDSKCFTAGTNTVTFNDGSKKLSIAVNFEKSTVDQ 180

Qy       181 SAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLND 240
Db       181 SGYLTTSRFMPSLNKIATLYVAPQCENGYTSGTMGFSTSYGDVAIDCSNVHIGISKGVND 240

Qy       241 WNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db       241 WNHPVTSESFSYTKSCSSFGISITYQNVPAGYRPFIDAYISPSDNNQYQLSYKNDYTCVD 300

Qy       301 SRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360
Db       301 DYWQHAPFTLKWTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKTIEILQ 360

Qy       361 PIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420
Db       361 PIPTTTITTSYVGVTTSYSTKTAPIGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNP 420

Qy       421 TDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 480
Db       421 TDSIDTVVVQVPSPNPTTTTTQFWSESFTSTTTITNSLKGTDSVIVREPHNPTVTTTEFW 480

Qy       481 SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 540
Db       481 SESFATTETITSKPEGTDSVIVREPHNPTVTTTEFWSESYATTETITNGPEGTDSVIVRE 540

Qy       541 PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTA 600
Db       541 PHNPTVTTTKFWSESYATTETITNKPEGTDSVIVKEPYNPTVTTTEFWSESYATTETITN 600

Qy       601 PPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTT--- 657
Db       601 GPEGTDSVIVREPHNPTVTTTEFWSESYATTETITTGPLGTDSIVIHDPLEESSSTTAIE 660

Qy       658 ----------EYWSQSYATTTTITAPPGETDTVLIREPPNHTVT---------------- 691
Db       661 SSDSNISSSAQESSSSVEQSSSIVGLSSSSDIPLSSDMPSSSSTGLTSSESSTVSSYDSD 720

Qy       692 ----------------------TTEYWSQS----YATTTTVTAPPGETDTVLIREPPNHT 725
Db       721 SSSSIESSTLSSSDRCSSSISDTTSFWDSSSSDLESTSITWSSSIDAQSSHLVQSVSNSI 780

Qy       726 VTTTEYWSQSYATTTTVTAPPGGTDTVIIREP------------PNPTVTTTEYW----- 768
Db       781 STSQELSSSSSEESSTF-----ATDALVSSDASSILSSDTSSYYPSSTISSSDDFPHTIA 835

Qy       769 ------SQSFATTT--------TVTAPPGGT-DTVIIYESMSSSKISTSSNDI------- 806
Db       836 GESDSLSISFITSTVEISSDSVSLTSDPASSFDSSSSLNSDSSSSPSSDQSDILTSSSFS 895

Qy       807 TSIIPSFS----------RPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTES 856
Db       896 TLVVPSFSLSSSSSLSLTYPHYVNSTTYHASESESSSVASPSMASESAN--DDTYTLSES 953

Qy       857 ETTTELICSDGKE---CSRLSSSSGIVTN-PDSNESSIVTSTVPTASTMSDSLSSTDG-I 911
Db       954 TDTTSSIGTDSSTVTFCRRDNGDGCIVTGMPSSSIDSEQTSDVTTTSSFVASSTPTSAEQ 1013

Qy       912 SATSSDNV--SKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTS 969
```

```
Db       1014 SITDNPNIDSSQTSASSSTKSSVSVSDTVVNSILLSETS--TLSSDDSTS-SDTSISSTT 1070

Qy        970 NGD----NQSGTHDSQSTSTEIEIVTTS---STKVLPPVVSSNTDLTSEPTNTR------ 1016

Db       1071 NSDTGNINAGSSHTSTASIKESSIQKTGVTLSSSYLSTKLSSTSDITIELITTELITTEL 1130

Qy       1017 -----EQPTTLSTTSNSITE------DITTSQPTGDNGDNTSSTNPVPTVAT----STLA 1061

Db       1131 TTIEDNEPNTFTSTPSSHSEIFSSDNSVLSKQVDRESTIKTSPTTDVTTVSSLSVHSTEA 1190

Qy       1062 SA-----SEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTD 1116

Db       1191 STATLGENSFSNVASTPSNIATSLRSTSSSSNHATESSGTVKSEASAEAIPSPPTS--TD 1248

Qy       1117 VTTEPTDTREQPTTLSTTSKTN-----SELVATTQATN---ENGGKSPSTDLTSSLTTGT 1168

Db       1249 NRLSYSTEEAKGITYANSGSTNNLITESQVAAPTDSTSVLIENPVVTSTFDDNSSAAVDQ 1308

Qy       1169 SASTSANSELV----TSGSVTGGAVASASNDQSHSTSV----------------TNSNSI 1208

Db       1309 PSKTKSIEESIMNPDSTNETNNGFIATLSQAQVPSSSIHSELISTTTAKTTDASMNGDSA 1368

Qy       1209 VSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLF 1259

Db       1369 ASNSQPTTLIQQVATS--SYNQPLITTYAGSSSATKHPSWLLKFISVALFF 1417


RESULT 3
S51959
hypothetical protein YAL063c - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C;Accession: S51959
R;Bussey, H.; Kaback, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.; Fortin, N.; Hall, J.; Ouel
submitted to the EMBL Data Library, August 1994
A;Description: The sequence of chromosome 1 of Saccharomyces cerevisiae.
A;Reference number: S51956
A;Accession: S51959
A;Molecule type: DNA
A;Residues: 1-1367 <BUS>
A;Cross-references: EMBL:U12980; GSPDB:GN00001; MIPS:YAL063c
C;Genetics:
A;Gene: SGD:FLO9; MIPS:YAL063c
A;Cross-references: SGD:S0000059
A;Map position: 1L

  Query Match             14.5%;  Score 939;  DB 2;  Length 1367;
  Best Local Similarity   28.1%;  Pred. No. 2.9e-33;
  Matches  410;  Conservative  214;  Mismatches  469;  Indels  364;  Gaps   84;

Qy         28 NSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKYTTSQT----S 83

Db         50 DSSTYSNAAYMAY---GYASKT-----KLGSVGGQTDISIDYNIPCV---SSSGTFPCPQ 98

Qy         84 VDLTAD-GVK-YATCQ------FYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPIA-- 133

Db         99 EDLYGNWGCKGIGACSNNPIIAYWSTDLFGFYTTPT--------------NVTLEMTGY 143

Qy        134 FNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSL 193

Db        144 FLPPQTGSYT----------------------------FKFATVDDSAIL--------- 165

Qy        194 NKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDC----------SNIHIGITKGLNDWN- 242

Db        166 ----------------------SVGGSIAFECCAQEQPPITSTNFTI---NGIKPWNG 198

Qy        243 -YPVSSESFSYTKTCTSNGIQIKYQNVPA-GYRPF----------IDAYISATDVN- 286

Db        199 SPPDNITGTVYMYAGFYYPMKIVYSNAVAWGTLPISVTLPDGTTVSDDFEGYVYTFDNNL 258

Qy        287 ---QYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDA-------GSNGI-----VIVATT 331

Db        259 SQPNCTIPDPSNYTVSTTITTTEP----WTGTFTSTSTEMTTVTGTNGVPTDETVIVIRT 314

Qy        332 RTVTDSTTAVTTLPFNPSVDKTKT----------------IEILQP-IPTTTITT---- 369

Db        315 PT-TASTIITTTEPWNSTFTSTSTELTTVTGTNGVRTDETIIVIRTPTTATTAITTTEPW 373

Qy        370 --SYVGVTTSYLTKTA----PIGETATVIVDVPYHTTT--TVTSEWTGTIT------TTT 415

Db        374 NSTFTSTSTELTTVTGTNGLPTDET-IIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTV 432

Qy        416 TRTN--PTDSIDTVVVQVPLPNPT-VSTTEYWSQSFATT----TTVTAPPG-GTD--TVI 465

Db        433 TGTNGLPTDE-TIIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTVTGTNGLPTDETIIV 491

Qy        466 IREPPNHT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPPNPT--VTTTE 514

Db        492 IRTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTDETIIVIRTPTTATTAMTTTQ 551

Qy        515 YWSQSFATT----TTVTAPPG-GTDS--VIIREPPNPT--VTTTEYWSQSYATT----TT 561

Db        552 PWNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTEPWNSTFTSTSTEMTT 611

Qy        562 VTAPPG-GTDS--VIIREPPNHT--VTTTEYWSQSYATT----TTVTAPPG-GTD--TVI 609

Db        612 VTGTNGLPTDETIIVIRTPTTATTAITTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIV 671

Qy        610 IREPPNHT--VTTTEYWSQSFATT----TTVTGP---PSGTDTVIIREPPNPT--VTTTE 658

Db        672 IRTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTDETIIVIRTPTTATTAMTTTQ 731

Qy        659 YWSQSYATT----TTITAP---PGETDTVLIREPPNHTV--TTTEYWSQSYATT----TT 705

Db        732 PWNDTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTT 791

Qy        706 VTAPPGE-TD--TVLIREPPNH--TVTTTEYWSQSYATT----TTVTAPPG-GTD--TVI 753

Db        792 VTGTNGQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMTTITGTNGVPTDETVIV 851

Qy        754 IREPPNPTV--TTTEYWSQSFATT----TTVTAPPG--GTDTVIIYESMS---SSKISTS 802

Db        852 IRTPTSEGLISTTTEPWTGTFTSTSTEMTTITGTNGQPTDETVIIVKTPTTAISSSLSSS 911

Qy        803 SNDITSIIPS---FSRPHYVNSTTSDLSTFESSSMNTPTSISSDGM---LLSSTTLVTES 856

Db        912 SGQITSFITSARPIITPFYPSNGTSVISSSVISSSDTSSLVISSSVTSSLVTSSPVISSS 971

Qy        857 ETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSS 916

Db        972 FISSPVISSTTTSASILSESS------------KSSVIPTSSSTSGSSESETGSASSAS 1018

Qy        917 DNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSG 976

Db       1019 ---SSSSISSESPKSTYSSSSLP-PVTSATTSQEITSSLPPVTTTKT------------ 1061

Qy        977 THDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPT---NTREQPTTLSTTSNSITEDI 1033

Db       1062 -----SEQTTLVTVTSCESHVCTESISSAIVSTATVTVSGATTEYTTWCPISTTEITKQT 1116

Qy       1034 T-TSQPTGDNGDNTSSTNPVPTVATSTLASASEED--NKSGSHESASTSLKPSMGENSGL 1090

Db       1117 TETTKQTKGTTEQTTETTKQTTVVT---ISSCESDVCSKTASPAIVSTSTATINGVTTEY 1173

Qy       1091 TTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSELVAT-----T 1145

Db       1174 TTWCPISTTESKQQTTLVTVTSCGSGVCSETT----SPAIVSTATATVNDVVTVYSTWRP 1229

Qy       1146 QATNENGGKSPSTDLTSSLTTGTSAST------SANSELVTSGSVT-----GGAVASASN 1194

Db       1230 QTTNEQSVSSKMNSATSETTTNTGAAETTTSTGAAETKTVVTSSISRFNHAETQTASATD 1289

Qy       1195 DQSHSTSV-----TNSNSIVSNTPQTTLSQQVTSS------SPSTNTFIASTYDGSGSII 1243

Db       1290 VIGHSSSVVSVSETGNTKSLTSSGLSTMSQQPRSTPASSMVGSSTASLEISTYAGSANSL 1349
```

```
Qy       1244 QHSTWLYGLITLLSLFI 1260
              : |   |  | | |
Db       1350 LAGSGLSVFIASLLLAI 1366


RESULT 4
S53465
flocculation protein FLO1 precursor - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YAR050w
C;Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 29-Oct-1999.
C;Accession: S53465; S43543; S57851; S31230; S57349
R;Bussey, H.; Keng, T.; Storms, R.K.; Vo, D.; Zhong, W.; Fortin, N.; Barton, A.B.; Kabac
submitted to the EMBL Data Library, February 1994
A;Description: Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of the 5
A;Reference number: S53458
A;Accession: S53465
A;Molecule type: DNA
A;Residues: 1-1537 <BUS>
A;Cross-references: EMBL:L28920; NID:g1616966; PIDN:AAC09499.1; PID:g694125; MIPS:YAR05C
R;Watari, J.; Takata, Y.; Ogawa, M.; Sahara, H.; Koshino, S.; Onnela, M.L.; Airaksinen,
Yeast 10, 211-225, 1994
A;Title: Molecular cloning and analysis of the yeast flocculation gene FLO1.
A;Reference number: S43543; MUID:94262325; PMID:8203162
A;Accession: S43543
A;Molecule type: DNA
A;Residues: 1-428,'M',430-473,'M',475-518,'M',520-549,'T',551-608,'L',610-636,'M',638-69
A;Cross-references: EMBL:X78160
R;Onnela, M.
submitted to the EMBL Data Library, September 1994
A;Reference number: S57851
A;Accession: S57851
A;Molecule type: DNA
A;Residues: 1-428,'M',430-463,'D',465-473,'M',475-518,'M',520-549,'T',551-608,'L',610-63
A;Cross-references: EMBL:X78160; NID:g535933; PIDN:CAA55024.1; PID:g535934
R;Teunissen, A.W.R.H.; Holub, E.; van der Hucht, J.; van den Berg, J.A.; Steensma, H.Y.
Yeast 9, 423-427, 1993
A;Title: Sequence of the open reading frame of the FLO1 gene from Saccharomyces cerevisi
A;Reference number: S31230; MUID:93289821; PMID:8511970
A;Accession: S31230
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-296,927-1516,'TAYWPVVV' <TEU>
R;Bidard, F.; Bony, M.; Blondin, B.; Dequin, S.; Barre, P.
Yeast 11, 809-822, 1995
A;Title: The Saccharomyces cerevisiae FLO1 flocculation gene encodes for a cell surface
A;Reference number: S57349; MUID:96090130; PMID:7483845
A;Accession: S57349
A;Molecule type: DNA
A;Residues: 1243-1274;1308-1339;1359-1390 <BID>
C;Genetics:
A;Gene: SGD:FLO1
A;Cross-references: SGD:S0000084; MIPS:YAR050w
A;Map position: 1R
C;Keywords: duplication; glycoprotein; transmembrane protein
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-1537/Product: flocculation protein FLO1 #status predicted <MAT>
F;953-997/Domain: repeat A2 <RA2>
F;998-1042/Domain: repeat A3 <RA3>
F;1043-1081/Domain: repeat A4 <RA4>
F;1226-1276/Domain: repeat B1 <RB1>
F;1277-1284/Domain: repeat B2 (partial) #status atypical <RB2>
F;1291-1341/Domain: repeat B3 <RB3>
F;1342-1392/Domain: repeat B4 <RB4>
F;1408-1416/Domain: repeat C1 <RC1>
F;1417-1425/Domain: repeat C2 <RC2>
F;1426-1434/Domain: repeat C3 <RC3>
F;135,187,262,1114/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match           14.5%;  Score 939;  DB 2;  Length 1537;
  Best Local Similarity  25.8%;  Pred. No. 3.4e-33;
  Matches  414;  Conservative  246;  Mismatches  520;  Indels  426;  Gaps   84;

Qy          2 LQQFTLLFLYLSIASAKT---------ITGVFDSF------NSLTWSNAANYAF------ 40
              |  |||| |  |:||  |          :|:  :|      :| |:||||  |:
Db         10 LAVFTLLAL-TSVASGATEACLPAGQRKSGMNINFYQYSLKDSSTYSNAAYMAYGYASKT 68

Qy         41 ----------------------------KGPGYPTW----------NAVLGWSLD--G 58
                                          :  |  |          :  || |  |
Db         69 KLGSVGGQTDISIDYNIPCVSSSGTFPCPQEDSYGNWGCKGMGACSNSQGIAYWSTDLFG 128

Qy         59 TSANPGDTFTLNM----------PCVFKYTTSQ----------------------TSVD 85
                | :  || |             ||: |                          || :
Db        129 FYTTPTNV-TLEMTGYFLPPQTGSYTFKFATVDDSAILSVGGATAFNCCAQQQPPITSTN 187

Qy         86 LTADGVK----------YATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGT----VTLP 131
                | ||:|           |   |:|  :             : |:  ::||    ||||
Db        188 FTIDGIKPWGGSLPPNIEGTVYMYAGYYYPM---------KVVYSNAVSWGTLPISVTLP 238

Qy        132 IAFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTV-DPSAYLYASRVM 190
                 ||  | | |           :|:|       |:     || ||| |   |
Db        239 -----DGTTVSDDFEG--------YVYSFDD-------DLSQSNCTVPDPSNYA-VSTTT 277

Qy        191 PSLNKVTTLFVAPQCENGYTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESF 250
               :    |  | :  |    :|| | :    : |         |      ||   :| |
Db        278 TTTEPWTGTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTTASTIITTTEPWNSTFTSTST 337

Qy        251 SYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTL 310
               |   :||::     :       |    :||     |  : : :|   : |
Db        338 ELTTVTGTNGVRTDETII------VIRTPTTATTAITTTEPWNSTFTSTSTEL------- 384

Qy        311 RWTGYKNSDAGSNGI---VIVATTRTVTDSTTAVTTL-PFNPSVDKTKT---------- 355
                    :  |:||:     :   || | :|||:||  |:| :   | |
Db        385 ------TTVTGTNGLPTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTVTGTNGL 438

Qy        356 ------IEILQPIPTTTITT-------SYVGVTTSYLTKTA----PIGETATVIVDVPYH 398
                    | |  |   ||  |       ::   :|   | |     |  ||  :::  |
Db        439 PTDETIIVIRTPTTATTAMTTTQPWNDTFTSTSTELTTVTGTNGLPTDET-IIVIRTPTT 497

Qy        399 TTT--TVTSEWTGTIT------TTTTRTN--PTDSIDTVVVQVPLPNPTVSTT-----EY 443
               ||  | |  |  | |      || | ||  |||  :|::|   :  || :||     :
Db        498 ATTAMTTTQPWNDTFTSTSTEITTVTGTNGLPTD--ETIIV---IRTPTTATTAMTTPQP 552

Qy        444 WSQSFATT----TTVTAPPG-GTD--TVIIREPPNHT--VTTTEYWSQSFATT----TTV 490
               |: :| :|    ||||   |  ||   ::|| |   |  :|||| |: :| :|    |||
Db        553 WNDTFTSTSTEMTTVTGTNGLPTDETIIVIRTPTTATTAITTTEPWNSTFTSTSTEMTTV 612

Qy        491 TAPPG-GTDS--VIIREPPNPT--VTTTEYWSQSFATT----TTVTAPPG-GTDS--VII 538
               |   |  ||   ::|| |   |  :|||: |: :| :|    ||||   |  ||   ::|
Db        613 TGTNGLPTDETIIVIRTPTTATTAITTTQPWNDTFTSTSTEMTTVTGTNGLPTDETIIVI 672

Qy        539 REPPNPT--VTTTEYWSQSYATT----TTVTAPPG-GTDS--VIIREPPNHT--VTTTEY 587
               | |   | :|||: |: :: :|    ||||   |  ||   ::|| |   |  :|||:
Db        673 RTPTTATTAMTTTQPWNDTFTSTSTEITTVTGTTGLPTDETIIVIRTPTTATTAMTTTQP 732

Qy        588 WSQSYATT----TTVTAPPG-GTD--TVIIREPPNHTV--TTTEYWSQSFATT----TTV 634
               |: :: :|    ||||   |  ||   ::|| | :  :  |||| |: :| :|    |||
Db        733 WNDTFTSTSTEMTTVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTV 792

Qy        635 T---GPPSGTDTVIIREPPNP--TVTTTEYWSQSYATT----TTITAP---PGETDTVLI 682
               |   | |:    ::|| | :     |||| |: :: :|    ||||      | :   ::|
Db        793 TGTNGQPTDETVIVIRTPTSEGLVTTTTEPWTGTFTSTSTEMTTITGTNGVPTDETVIVI 852

Qy        683 REPPNHTV--TTTEYWSQSYATT----TTVTAPPGE-TD--TVLIREPPNHTV--TTTEY 731
               | | :  :  |||| |: :: :|    ||:|   |: ||   ::|| | :  :  ||||
Db        853 RTPTSEGLISTTTEPWTGTFTSTSTEMTTITGTNGQPTDETVIVIRTPTSEGLISTTTEP 912

Qy        732 WSQSYATTTTVTAPPGGTD-------TVIIREPPNPTV--TTTEYWSQSFATTTTVTAPP 782
               |: :: :|:|      ||:       ::|| | :  :  |||| |: :| :|:|
Db        913 WTGTFTSTSTEMTHVTGTNGVPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEVTTI 972

Qy        783 GGT------DTVIIYESMSSSKI-----------STSSNDITSIIPSFSRP-------- 816
               ||      :|||:  : :|  :           :::| ::|::   :  :|
Db        973 TGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVI 1032
```

```
Qy        817 ------HYVNSTTSD-LSTFESSSMNTPTSISSDGMLLSSTTLVTESETT----TELICS 865
Db       1033 RTPTSEGLVTTTTEPWTGTFTSTSTEMSTVTGTNGLPTDETVIVVKTPTTAISSSLSSSS 1092

Qy        866 DGKECSRLSSSSGIVTNP--DSNESSIVTSTVPTASTMSDSLSSTDGISATSSDNVSKSG 923
Db       1093 SGQITSSITSSRPIIT-PFYPSNGTSVISSSVISSSVTSSLFTSSPVISSSVISSSTTTS 1151

Qy        924 VSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQST 983
Db       1152 TSIFSESSKSSV--IPTSSSTSGSSESETSSAGSVSSS----SFISSESSKSPTYSSSS- 1204

Qy        984 STEIEIVTTSST-----KVLPPVVSSNTDLTSEPTNTREQPTTLSTTS-------NSITE 1031
Db       1205 ---LPLVTSATTSQETASSLPPATT---------TKTSEQTTLVTVTSCESHVCTESISP 1252

Qy       1032 DITTSQPTGDNGDNTSSTN--PVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSG 1089
Db       1253 AIVSTATVTVSGVTTEYTTWCPISTTETTKQTKGTTEQTTETTKQTTVVTISSCESDVCS 1312

Qy       1090 LTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTT-------SKTNSELV 1142
Db       1313 KTASPAIVSTSTATINGVTTEYTTWCPIST--TESRQQTTLVTVTSCESGVCSETASPAI 1370

Qy       1143 ATT----------------QATNENGGKSPSTDLTSSLTTGT-SASTSANSELVTSGSV 1184
Db       1371 VSTATATVNDVVTVYPTWRPQTANEESVSSKMNSATGETTTNTLAAETTTNTVAAETITN 1430

Qy       1185 TGGA------------------VASASNDQSHSTSV-----TNSNSIVSNTPQTTLSQQ 1220
Db       1431 TGAAETKTVVTSSLSRSNHAETQTASATDVIGHSSSVVSVSETGNTKSLTSSGLSTMSQQ 1490

Qy       1221 VTSSSP------STNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db       1491 PRSTPASSMVGYSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1536
```

RESULT 5
S48992
flocculation protein homolog YHR211w - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 19-Apr-2002
C;Accession: S48992
R;Macri, C.
submitted to the EMBL Data Library, February 1994
A;Description: The sequence of S. cerevisiae cosmid 9177.
A;Reference number: S46671
A;Accession: S48992
A;Molecule type: DNA
A;Residues: 1-1075 <MAC>
A;Cross-references: EMBL:U00029; NID:g551322; PID:g458919; GSPDB:GN00008; MIPS:YHR211w
C;Genetics:
A;Gene: SGD:FLO5; MIPS:YHR211w
A;Cross-references: SGD:S0001254
A;Map position: 8R

Query Match 10.9%; Score 711; DB 2; Length 1075;
Best Local Similarity 25.9%; Pred. No. 1.5e-23;
Matches 324; Conservative 192; Mismatches 393; Indels 342; Gaps 66;

```
Qy        153 AGTNTVTFNDGDKDISIDVEF------EKSTVDPSAYL---YASRV-------------- 189
Db         23 SGATEACLPAGQRKSGMNINFYQYSLKDSSTYSNAAYMAYGYASKTKLGSVGGQTDISID 82

Qy        190 --MPSLNKVTTLFVAPQCENGYTS-GTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVS 246
Db         83 YNIPCVSSSGT-FPCPQ-EDSYGNWGCKGMGA--------CSN-----SQGIAYW----S 123

Qy        247 SESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSR---L 303
Db        124 TDLFGFYTTPTNVTLEMTGYFLPPQTGSYTFSFATVDDSAILSVGGSIAFECCAQEQPPI 183

Qy        304 QSKPFTLRWTGYKNSDAGSNGIVIVATT----------RTVTDSTTAVTTLPFNPSVDKT 353

Db        184 TSTNFTI--NGIKPWD-GSLPDNITGTVYMYAGYYYPLKVVYSNAVSWGTLPI--SVE-- 236

Qy        354 KTIEILQPIPT-TTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTT---TTVTSEWTG 409
Db        237 --------LPDGTTVSDNFEGYVYSF---DDDLSQSNCTIPDPSIHTTSTITTTTEPWTG 285

Qy        410 TIT------TTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDT 463
Db        286 TFTSTSTEMTTITDTNGQLTDETVIV---IRTPT------------TASTIT-------- 322

Qy        464 VIIREPPNHTVTTTEYWSQSFATT----TTVTAPPG-GTDS--VIIREPPNP--TVTTTE 514
Db        323 -----------TTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLITTTTE 371

Qy        515 YWSQSFATT----TTVTAPPG-GTDS--VIIREPPNP--TVTTTEYWSQSYATTTTVTAP 565
Db        372 PWTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEVTT 431

Qy        566 PGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNH--TVTTTEY 623
Db        432 ITGTNG----QPTDETV------------------------IVIRTPTSEGLITTTTEP 462

Qy        624 WSQSFATT----TTVT---GPPSGTDTVIIREPPNPTV--TTTEYWSQSYATT----TTI 670
Db        463 WTGTFTSTSTEMTTVTGTNGQPTDETVIVIRTPTSEGLISTTTEPWTGTFTSTSTEVTTI 522

Qy        671 TAPPGE-TD--TVLIREPPNH--TVTTTEYWSQSYATT----TTVTAPPGE-TD--TVLI 718
Db        523 TGTNGQPTDETVIVIRTPTSEGLITTTTEPWTGTFTSTSTEMTTVTGTNGQPTDETVIVI 582

Qy        719 REPPNH--TVTTTEYWSQSYATTTTVTAPPGGT------DTVIIREPPNPTVTTTEYWSQ 770
Db        583 RTPTSEGLITRTTEPWTGTFTSTSTEVTTITGTNGQPTDETVIVIRTPT-TAISSSLSSS 641

Qy        771 SFATTTTVTA---------PPGGTDTV---IIYESMSSSKISTSSNDITSIIPSFSRPHY 818
Db        642 SGQITSSITSSRPIITPFYPSNGTSVISSSVISSSVTSSLVTSSSFISSSVISS------ 695

Qy        819 VNSTTSDLSTFESSSMNT--PTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSS 876
Db        696 --STTTSTSIFSESSTSSVIPTSSSTSGSSESKTSSASSSSSSSSISSESPKSPTNSSSS 753

Qy        877 SGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATS------SDNVSKSGVSVTTET 930
Db        754 LPPVTSATTGQET--ASSLPPATTTKTS-EQTTLVTVTSCESHVCTESISSAIVSTATVT 810

Qy        931 --SVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTHDSQSTSTEIE 988
Db        811 VSGVTTEYTTWCPISTTETT----------------------KQTKGTTEQTKGTTEQT 847

Qy        989 IVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSS 1048
Db        848 TETTKQTTVV-TISSCESDICSKTAS----PAIVSTSTATI-----------NGVTTEY 890

Qy       1049 TNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAP- 1107
Db        891 TTWCPISTTES------------------------KQQTTLVTVTSCESGVCSETTSPA 925

Qy       1108 --SPAVSSGTDVTT-----EPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDL 1160
Db        926 IVSTATATVNDVVTVYPTWRPQTTNEQ----SVSSKMNS---ATSETTTNTGAAETKTAV 978

Qy       1161 TSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSV-----TNSNSIVSNTPQT 1215
Db        979 TSSLSRFNHAETQ--------------TASATDVIGHSSSVVSVSETGNTMSLTSSGLS 1023

Qy       1216 TLSQQVTSS------SPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI 1260
Db       1024 TMSQQPRSTPASSMVGSSTASLEISTYAGSANSLLAGSGLSVFIASLLLAI 1074
```

RESULT 6

```
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N;Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR019c
C;Species: Saccharomyces cerevisiae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 12-Nov-1999
C;Accession: S48478; A26877; B26877; S27281; JC6123
R;Rowley, K.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48478
A;Accession: S48478
A;Molecule type: DNA
A;Residues: 1-1367 <ROW>
A;Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN00009; MIP
R;Yamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A;Title: Gene fusion is a possible mechanism underlying the evolution of STA1.
A;Reference number: A91831; MUID:87194600; PMID:3106330
A;Accession: A26877
A;Molecule type: DNA
A;Residues: 1-242 <YAM>
A;Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525
A;Accession: B26877
A;Molecule type: DNA
A;Residues: 762-1331 <YA2>
A;Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526
R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
A;Reference number: S27281; MUID:89031230; PMID:3141213
A;Accession: S27281
A;Molecule type: DNA
A;Residues: 1-31 <PAR>
A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
A;Reference number: JC6123; MUID:96323237; PMID:8710886
A;Accession: JC6123
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1367 <LAM>
A;Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387
C;Genetics:
A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458
A;Cross-references: MIPS:YIR019c; SGD:S0001458
A;Map position: 9R
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein
F;5-21/Domain: transmembrane #status predicted <TM1>
F;1350-1366/Domain: transmembrane #status predicted <TM2>

  Query Match            10.4%;  Score 675.5;  DB 1;  Length 1367;
  Best Local Similarity  26.2%;  Pred. No. 7.1e-22;
  Matches  378;  Conservative  177;  Mismatches  535;  Indels  353;  Gaps   65;

Qy          1 MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTS 60
              | : | | :| ||:            |||              |:||     | | :|||
Db          1 MQRPFLLAYLVLSLL----------FNSAL----------GFPTALVPRG-SSEGTS 36

Qy         61 AN------PGDTFTLNM--PCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTC 112
               |      |   | :|   : :||   |||    |   |       |:|
Db         37 CNSIVNGCPNLDFNWHMDQQNIMQYTLDVTSVSWVQDNT-YQITIHVKGKE--------- 86

Qy        113 TVNDALKS--SIKAFGTVTLPIAFNVGGTGSSTDLEDSKC-FTAGTNTVTFN-DGDKDI- 167
                | ||   |:|  |          : |   :| | |:   |||     ||     :|:
Db         87 --NIDLKYLWSLKIIGVTGPKGTVQLYGYNENTYLIDNPTDFTA-----TFEVYATQDVN 139

Qy        168 -------SIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSSSN 220
                    :  ::||        ||  |               | |  : :  ||  |  |
Db        140 SCQVWMPNFQIQFE--------YLQGS--------------AAQYASSWQWGTTSFDLST 177

Qy        221 GDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFIDAYI 280
              |     |:|      :| :  ::|     | :   | :|

Db        178 G-----CNNYD---NQGHSQTDFP----GFYWNIDCDNN--------------------C 205

Qy        281 SATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTGYKNSDAGSNGIVIVATTRTVTDSTTA 340
                |  :  | : :   |  |  :|  |        :| : |:      |: |   ||::
Db        206 GGTKSSTTTSSTSESSTTTSSTSESSTTT-------SSTSESS----TTTSSTSESSTSS 254

Qy        341 VTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYLTKTAPIGETATVIVDVPYHTT 400
               || |  |:     |    :  ||   |||         ||  |           |:| |
Db        255 STTAPATPT-----TTSCTKEKPTPPTTTS--------CTKEKP---.-------TPPHHDT 292

Qy        401 TTVTSEWTGTITTTTTRTN---PTDSIDTV-VVQVPLPNPTVSTTEYWSQ--SFATTTTV 454
              |  | : | |  | | :|    || |  |       |:| |: |||  |   : :|| :
Db        293 TPCTKKKTTTSKTCTKKTTTPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESS 352

Qy        455 TAP---PGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVT 511
              :||   |  : |     |   : | :    : :|| : :||          |  | |: :
Db        353 SAPVPTPSSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAP----------VPTPSSS 401

Qy        512 TTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTT-VTAPPGGTD 570
              |||  |    ::|| ::    | |          | :|||  |    ::||  :: |  |
Db        402 TTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVTSSTTESSSAPVPTP 461

Qy        571 SVIIREPPNHTVT--TTEYWS-----QSYATTTTVTAPPGGTDTVIIREP-PNHTVTTTE 622
              |    |  :  ||  |||  |       | :|| : :||   : |     | |  : :|||
Db        462 SSSTTESSSAPVTSSTTESSSAPVPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTE 521

Qy        623 YWS-----QSFATTTTVTGPPSGTDTVIIREP-PNPTVTTTEYWSQSYATTTTITAPPGE 676
                |      | :|| : : | : :|     | | |: :|||      :::| :|:   |
Db        522 SSSAPAPTPSSSTTESSSAPVTSSTTESSSAPVPTPSSSTTE------SSSTPVTSSTTE 575

Qy        677 TDTVLIREPPNHTVTTTEYWSQSYAT----TTTVTAPPGETDTVLIREPPNHTVT--TTE 730
              : :  :   |  : :|||  |     |    ||  :: | |  | :    |  :  ||  |||
Db        576 SSSAPV---PTPSSSTTESSSAPVPTPSSSTTESSSAPAPTPSSSTTESSSAPVTSSTTE 632

Qy        731 YWS-----QSYATTTTVTAP---PGGTDTVIIREP-PNPTVTTTEYWSQSFATTTTVTAP 781
                |      | :|| : :||   | : |     | | |: :|||  | |   |:: |
Db        633 SSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTE--SSSAPVTSSTTES 690

Qy        782 PGGTDTVIIYESMSS-----SKISTSSNDITSIIPSFSR--------PHYVNSTTSD--- 825
                  |   || |:     |  :| |:      || |         |   :|||
Db        691 SSAPVTSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSSA 750

Qy        826 ---LSTFESSSMNTPTSISSDGMLLSSTTLVTESETTTELICSDGKECSRLSSSSGIVTN 882
                 || |||   ||  ||      ||  : | | :|||   :     |  :: | :
Db        751 PVTSSTTESSSAPVPTP-SSSTTESSSAPVPTPSSSTTESSSAPVPTPSSSTTESSVAPV 809

Qy        883 PDSNESSIVTSTVPTASTMSDSL-SSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNP 941
              |  : || :||: |:::  | |  ||:  :   ||     |   |:: |: :::   | |
Db        810 PTPSSSSNITSSAPSSTPFSSSTESSSVPVPTPSSSTTESSSAPVSSSTTESSVAPVPTP 869

Qy        942 LSSSVTSLTQLSSIPSVSESES-----KVTFTSNGDNQSGTHDSQSTSTEIEIV------ 990
               |||  : :  ||||  | :||     || :|:    | |  | |::||  ||
Db        870 SSSSNITSSAPSSIPFSSTTESFSTGTTVTPSSSKYPGSQTETSVSSTTETTIVPTKTTT 929

Qy        991 ---TTSSTKVLPPVVSSNTDLTSEPTN-------TREQPTTLST--TSNSITEDITTSQP 1038
                  | |:| :   | |: |:   | |:       |   ||| :|  |::| |   ||
Db        930 SVTTPSTTTITTTVCSTGTNSAGETTSGCSPKTVTTTVPTTTTTSVTTSSTTTITTTVCS 989

Qy       1039 TGDN--GDNTSSTNPVPTVATSTLASASEEDNKSGSHESA-----------------ST 1078
              || |  |: ||  :|  |: |:   | |  :  | |  ::                 ||
Db        990 TGTNSAGETTSGCSP-KTITTTVPCSTSPSETASESTTTSPTTPVTTVVSTTVVTTEYST 1048

Qy       1079 SLKPSMGENSGLTTS--TEIEATTTSPTEAPSPAV---------------SSGTD---- 1116
              | ||  ||   :||:  |:   ||   | ||:|:|              |:||:
Db       1049 STKPG-GE---ITTTFVTKNIPTTYLTTIAPTPSVTTVTNFTPTTITTTVCSTGTNSAGE 1104

Qy       1117 ----------VTTEPTDT-----REQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLT 1161
                       || |  |       : ||| ||: | : :| |  :|  |     :|  |
Db       1105 TTSGCSPKTVTTTVPCSTGTGEYTTEATTLVTTAVTTT-VVTTESSTGTNSAGKTTTGYT 1163

Qy       1162 SSLTTGTSASTSANSELVTSG------SVTGGAVASASNDQSHSTSVTNSNSIVSN---- 1211
              :    |  :| | |  ||        ::|    ::|:|    :||| :: :|||:
Db       1164 TKSVPTTYVTTLAPSAPVTPATNAVPTTITTTECSAATNAAGETTSVCSAKTIVSSASAG 1223
```

```
Qy       1212 --------TPQTT---LSQQVTSSSPSTN----------------TFIASTYDGSGSIIQ 1244
                      || ||    :   | ||  ||                |:| :   ||
Db       1224 ENTAPSATTPVTTAIPTTVITTESSVGTNSAGETTTGYTTKSIPTTYITTLIPGSNGAKN 1283

Qy       1245 HST 1247
              : |
Db       1284 YET 1286


RESULT 7
T33369
hypothetical protein H02F09.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000
C;Accession: T33369
R;Geisel, C.; Harmon, G.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid H02F09.
A;Reference number: Z21330
A;Accession: T33369
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1275 <GEI>
A;Cross-references: EMBL:AF077538; PIDN:AAC64622.1; GSPDB:GN00028; CESP:H02F09.3
A;Experimental source: strain Bristol N2; clone H02F09
C;Genetics:
A;Gene: CESP:H02F09.3
A;Map position: X
A;Introns: 42/1; 78/3; 106/3; 135/2; 161/2; 224/1
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

  Query Match           10.0%;  Score 649.5;  DB 2;  Length 1275;
  Best Local Similarity  26.8%;  Pred. No. 8.7e-21;
  Matches  326;  Conservative  151;  Mismatches  449;  Indels  289;  Gaps   53;

Qy        213 TMGFSSSNG----------DVAIDCS------NIHIGITKGLNDWNYPVSSESFSYTKT 255
              |:   :||           |   |||       :   :|    | |: ::: | : |||
Db         49 TVSLCNSNSEPQISSLRFKRDANTDCSFNANKNQLRNFLTSSAEDANFILTTYSGTSTKT 108

Qy        256 ---------------CTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
                            :|  :    | |   || | ||:  :|   ||
Db        109 EPLTKQLALDQLEAIKADGSSAVKQSSAIESYDP--SAY-SNTDLVFFTPCQTN------ 159

Qy        301 SRLQSKPFTLRWTG----YKNSDAGSNGI-----VIVATTRTVTDSTT----AVTTLPFN 347
                  . : |    ||     | :     |||: :   ||  :    |   :|
Db        160 ----------YDGIEDDIKNVQTAINKVITKTFVIVSLSLNSTDMNSRYGEAAHNIPTT 208

Qy        348 PSVDKTKTIEILQPIPTT---TITTSYVGVTTSYLTKTAP-----IGETATVIVDVPYHT 399
              |: | :  |  :  | ||    :||| :  ||: :|  ||     |  : | :| ||
Db        209 PTEDISNKINNILNIGTTQTPPVTTSTMATTTANVTSAAPNTTVTISTSPTTVVTVP--- 265

Qy        400 TTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPG 459
              :|  ||  |     ||| | |     | || ||    |: :|   |     :| || |
Db        266 STAQTSSTTTVTVPTTTVTGP-----TTVVTVPTTVVTIPSTVVTSPITTPSTVVTVP-- 318

Qy        460 GTDTVIIREPPNHTVTTTEYWSQSFATT---TTVTAPPGGTDSVIIREP-----PNPTVT 511
                ||:       || :|    |   |   | || |     | :: :|      :|||
Db        319 --STVV-------TVPSTAVTKPSTVVTAPSTVVTVP-----STVVTKPNTVVTSSPTVA 364

Qy        512 TTEYWSQSFATTTTVTAPPGGTDSVIIREPPN-PTVTTTEYWSQSYATTTTVTAPPGGTD 570
              ||         || || |     | ::  |    || ||   : |    | ||||
Db        365 TT--------PTTVVTTP-----STVVTVPSTVVTVPTTVVTNPS----TVVTAP----- 402

Qy        571 SVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTE-------- 622
              | ::      || ||   |:|   ||  |   ||:           |  |||
Db        403 STVV------TVPTTVMTSRSTVITTPTT---GGSSPSTAGTSLASTAVTTETSIGSSST 453

Qy        623 -------YWSQSFATTTTVTGPPSGTDTVIIREPPNPTV---------------TTTEY 659
                    | |  :| | :   :|  :   ::   ||:                 | :
Db        454 PLPSQSTSLSMSSLSTYTPSSSTAGATSPATQQSTKPTIGTSMSSGPTTVAPGASTESTV 513

Qy        660 WSQSYATTTTITAPPGETDTVLIREPPNHTVTT-TEYWSQSYATTTTVTAPPGETDTVLI 718
```

```
Db        514 LQSSTPSGTTVTLPSGSSTATAGTSPQASTVTTVTDISTVSGSTVTSQTA-----ESSLS 568

Qy        719 REPPN------HTVTTTEYWSQSY-------------ATTTTVTAPPGGTDTVIIREPPN 759
              | |         ||:|      :|             : :|  || || |::
Db        569 TESPTSAGSSISTVSTVSSQPSTYIPVSSASSIYSTLSGSTGSTASPGTTESSGSSTSGP 628

Qy        760 PTVTTTEYWSQSFATTTTVTAP------------PGGTD-TVIIYESMSSSKISTSSNDI 806
              |:: :   | |  | :|||               || |: ||    ::| | :|| |
Db        629 STISGS---SASTVTGSTVTEASTISGSTESSTIPGSTESTVSEASTVSGSSVSTVSGST 685

Qy        807 TSIIPSFSR-PHYVNSTTSDLSTFESS--SMNTPTSI----------------SSDGML 846
              |      |        || || ||   |  | | | |                  |:
Db        686 ESTSAGASTVSGSTGSTVSDSSTISDSTGSTNAPGSTESTVTGSSVSTVSGSTGSTGPST 745

Qy        847 LSSTTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESS---IVTSTVPTASTMSD 903
              :|::|  | :  :||   :||   |  : |:|   || | :||   | | :  : ||:|
Db        746 MSASTGSTNTPGSTESTITDGSTVSGSTGSTGSTNNPGSTDSSTTGISTVSGSSLSTISG 805

Qy        904 SLSSTDGISATSSDNVSKSGVSV--TTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSE- 960
              |  ||  :| :|   ||    |   :||::|:   |      ||| : |  ||: :||:
Db        806 STGST--VSGSSDMTVSTGSTSSPGSTESTVSGASTMSPSTGSSVETSTSGSSVSTVSQS 863

Qy        961 ----------SESKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLT 1009
                       ||| |:  |:    | :  | :|       :| ||      :|::|  |
Db        864 TSSSTTGQSTVSESSVSTVSSESTISQSTGSTTTGESTVFGSTGSTATGSSTMSASTGST 923

Qy       1010 SEPTNTREQPTTLSTTSNSITEDITTSQPTGD--NGDNTSSTNPVPTVATSTLASASEED 1067
              | :|       : | :::| : | |  ||       :| | : : ||  || :: : |
Db        924 DTPGSTES-----TITGSTVTGESTVSGSTGSTITEGSTISESTMTTVGVSTGSTITGES 978

Qy       1068 NKSGSHESASTSLKPSMGEN--SGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTR 1125
                |||  |  |      ||:  || | ||   :| ::||   : : |:|: || | |
Db        979 TVSGSTRSTVT------GESTVSGSTESTVSGSTESTPTVPSTVSGSTGSTVTGEST--- 1029

Qy       1126 EQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVT 1185
                : || | ::    ::|:| :   | | ||     :|: | :|||  |      :|:
Db       1030 ---VSGSTASTSSGSTGSSTEAGSTVSGSSAST-----VTSSTGSSTSGES------TVS 1075

Qy       1186 GGAVASASNDQSHSTSVTNSNSIVSNTPQT-TLSQQVTSSSPST---NTFIASTYDGSGS 1241
              |  |:: |   |  :::|  :::  :|  | |    |: || ||   ||   ||  | |
Db       1076 GSTVSTVSG--STGSTITGESTVSGSTESTVTAESTVSGSSVSTVSGNT--GSTITGE-S 1130

Qy       1242 IIQHSTWLYGLITLL 1256
                :  ||   |  |:|
Db       1131 TVSGSTGSTGESTIL 1145


RESULT 8
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T34434
R;Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid K06A9.
A;Reference number: Z21525
A;Accession: T34434
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2232 <GEI>
A;Cross-references: EMBL:U80846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A;Experimental source: strain Bristol N2; clone K06A9
C;Genetics:
A;Gene: CESP:K06A9.1a
A;Map position: X
A;Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 20

  Query Match           10.0%;  Score 646.5;  DB 2;  Length 2232;
  Best Local Similarity  23.8%;  Pred. No. 2.4e-20;
  Matches  355;  Conservative  221;  Mismatches  530;  Indels  384;  Gaps   61;
```

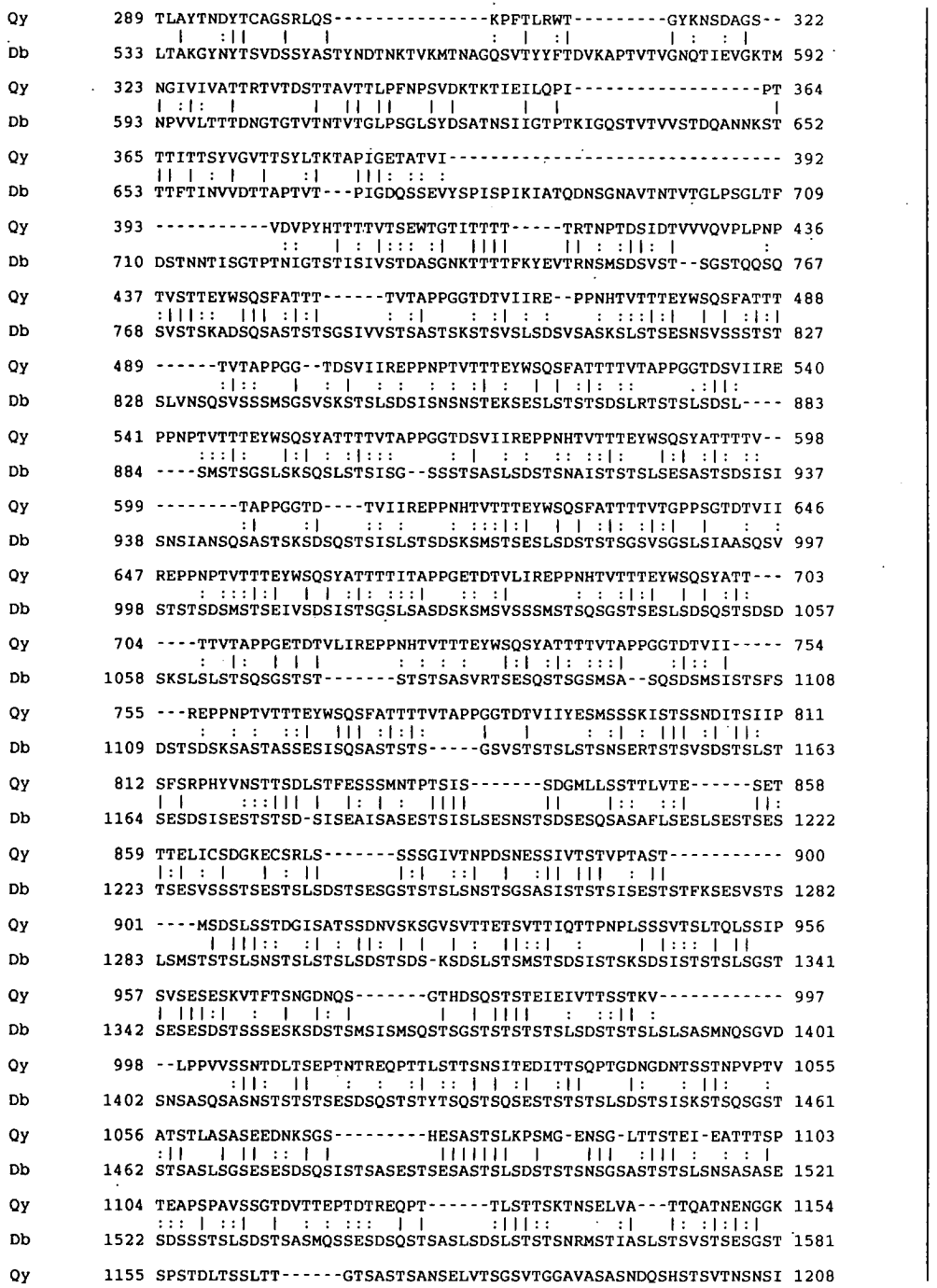
```
Qy        13 SIASAKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMP 72
Db       344 SSGTMSTISGSTGSTVTVVPGSSSTFASSTP------------IASSSSPGSTVTVAPG 390

Qy        73 CVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIKAFGTVTLPI 132
Db       391 SSSTYGSSTPSASSSSSG----TMSTNSG---STGSTVT-----VAPVSSSTFGSST-PI 437

Qy       133 AFNVGGTGSSTDLEDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPS 192
Db       438 A-SSSSSGSTVTVVSGSSSTYGSSTPSASSSSAGTA-------STISGSTGSTATIVPGS 489

Qy       193 LNKVTTLFVAPQCENGYTSGTMG-FSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFS 251
Db       490 SSSVGS---STQSASPSSPGTMSTVSGPTGST--------VTVVPGSSTSPAPSSSPNPS 538

Qy       252 YTKTCTSNGIQIKYQNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLR 311
Db       539 SSPASTGSTITI------SGSSSIIVSTVSGSTVSGST--GTSQSTLASS---------- 580

Qy       312 WTGYKNSDAGSNGIVIVATTRTVTDSTTAVTTLPFNPSVDKTKT-IEILQPIPTTTITTS 370
Db       581 -----TATPGSSSTVPSSSSPQPSSQSPAPNTGSTTPSQTSSQSPSPSMNPSSSTPTGSS 635

Qy       371 YVGVTTSYLTKTAPIGETATVIVDVPYHTTTTVTSEWT----GTITTTTTRTNP------ 420
Db       636 QSTITPEGSTASSPTGSTGST-----FSVATEVTSQSTVPSGSSLGTQSTNSSPSPSSLS 690

Qy       421 --TDSIDTV----------------VVQVPLPNPTVSTTEYWSQ-SFATT------TTVT 455
Db       691 PSTSGMSTLTSEPSPSSTQSSGAQSTLTTPSPNPSQSTSSLESSTSGATTSSGSAGTTMT 750

Qy       456 AP--------------------------------PG-----------GTDTVIIREPPNH 472
Db       751 SPSQSSSVGSSQGSTSPAASTTSGEMTSQGSTQTPGSSVSTSAAILTSTQQSVSTNSPGS 810

Qy       473 TVTTTEYWSQSFATTTTVTAPPGGTD------SVIIREP-------PNPT-------VTT 512
Db       811 TVTRPSTVSGSTSSGSTVTV--GSTEASTSGSSVASSSPAPSTSQNPNPSTSSGSSMITQ 868

Qy       513 TEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSV 572
Db       869 SPYPSQSTSPVESSTTPSPGS--------PGTTLTST---SPSPSQSTTIGSTQGSTS-- 915

Qy       573 IIREPPNHTVTTTEYWSQSY-----ATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQS 627
Db       916 -----PGISTTSEEMTSQGSTQTPGSTGSTVTQPSTVSDST--SSGSTVTVGSTEGSSSP 968

Qy       628 FATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREP-P 686
Db       969 IPSTSQNTNPSTSSGSSMSTQTPQSSQSTSPVESSTSGATSS-SGSPGTTLTSISPSPSP 1027

Qy       687 NHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTT-----TEYWSQSYATTTT 741
Db      1028 SSTIGS----SQGSTSPVVSTISQGSTET------PGSTGSTVTKPSTVSGSASSGSTAT 1077

Qy       742 V-----TAPPGGTDTVIIREPPNPTVTTTEYWS-------QSFATTTTVTAPPGGTDTVI 789
Db      1078 MGSTEASSTSGGSST-----SPNPSQSTSPSTSGATSSPGSSGTTLTSISPSPSQSSTIG 1132

Qy       790 IYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSS 849
Db      1133 SSQGSTSPVVSTTSGDMTS-QGSTQIPGSTGSTVTQPST-GSGSTSTSGEITSQG----S 1186

Qy       850 TTLVTESETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVT-STVPTAST-------- 900
Db      1187 TQTPRSSLSTSPAISTSTQQSVSTNSPGSTVTQPSTVRGSTSSGSTVTTGSTEGSSTSGS 1246

Qy       901 -MSDSLSSTDGISATS-SDNVSKSGVSVTTE---------TSVTTIQ-------TTPNPL 942
Db      1247 SSATSLSSSSPVPSTSQSPNPSTSGSSTPTPNPSQSTSPVVSTTTGEMTSHGSTQTPSTI 1306

Qy       943 SSSVTSLTQLSSIPSVSESESKVTFTSNGDNQSGTH----------------------D 979
```

```
Db      1307 GSTV---TQPSTVSGSNSSGSTVTIGSSEASTSGSSFKTSPSSISPVPTSSPIPSTTFAS 1363

Qy       980 SQSTSTEIEIVTTSSTKVLP----------------------------PVVSSNTDLTS 1010
Db      1364 STSGSTISDVSSVSTTSLAPLSSSLPSTVPSSTQSFSSTSEGSSKASSSPVPSQTSSTPT 1423

Qy      1011 EPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKS 1070
Db      1424 NPTGSTESSTLLSSTISGSTQHTTMSK--ASSGSTSPSTNS-QTGSTVTMGSSS----TS 1476

Qy      1071 GSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEP----TDTRE 1126
Db      1477 GVSTSSASSTQPQMSTSQGSSAGSTVASSTASPAASSTAPSSTGTMSSTSSGTVGSTISE 1536

Qy      1127 QPTTLSTTSKTNSEL---------VATTQATN-------ENGGKSPSTDLTSSLTTGTSA 1170
Db      1537 SSTTASASSQTGSTVTMGSSSTSGVSTSSASSTQPQMSTSQGSSAGST--VASSTAGLVS 1594

Qy      1171 STSANSELVTSGSVTGGAV----------ASASNDQ-------SHSTSVTNSNSIVSNTP 1213
Db      1595 TSTVPSSTGTMGSTSSGTVGSTISESSTTASASSQTGSTVTMGSSSTSGVSTSSASSTQP 1654

Qy      1214 QTTLSQ--------------QVTSSSPSTNTFIASTYDGS-GSIIQHST 1247
Db      1655 QMSTSQGSSAGSTVASSTTGLVSTSTVPSSTGTMGSTSSGTVGSTISESS 1704
```

RESULT 9
F90073
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: F90073
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: F90073
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2271 <KUR>
A;Cross-references: GB:BA000018; PID:g13702612; PIDN:BAB43752.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA2447

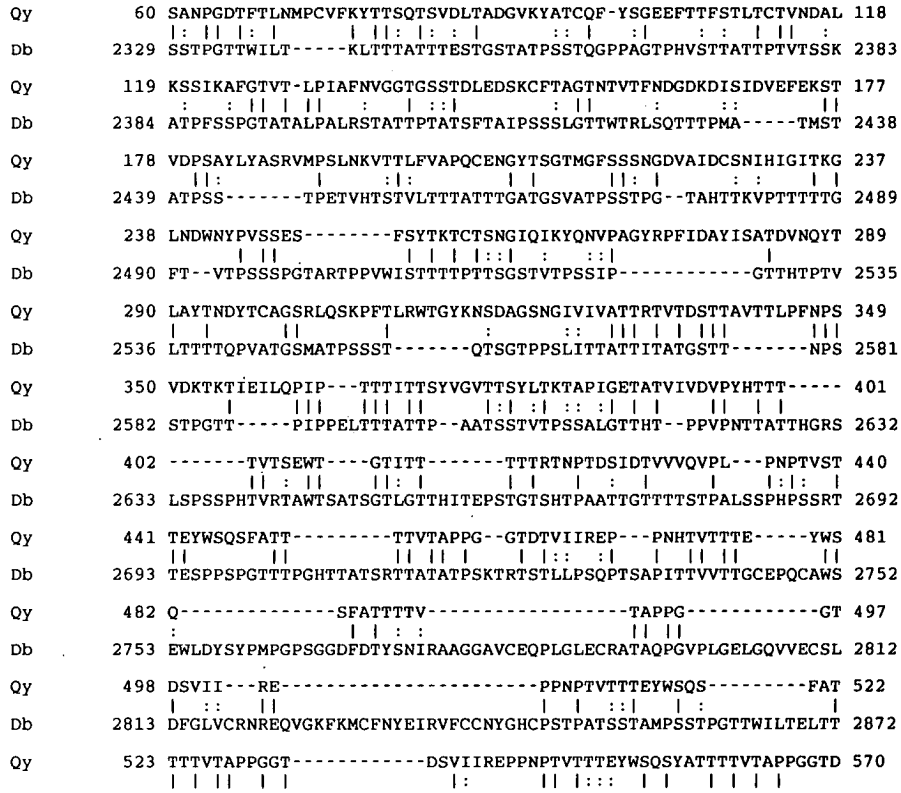
Query Match 9.8%; Score 637.5; DB 2; Length 2271;
Best Local Similarity 22.3%; Pred. No. 5.9e-20;
Matches 330; Conservative 280; Mismatches 521; Indels 347; Gaps 52;

```
Qy        17 AKTITGVFDSFNSLTWSNAANYAFKGPGYPTWNAVLGWSLDGTSANPGDTFTLNMPCVFK 76
Db       297 AITLGTRIDSNKSFHFSGKVNLGNKYEGHGNGGDGIGFAF-----SPG------------ 339

Qy        77 YTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCTVNDALKSSIK--------AFGTV 128
Db       340 -VLGETGLNGAAVGIGGLSNAF--GFKLDTYHN-TSKPNSAAKANADPSNVAGGGAFGAF 395

Qy       129 TLPIAFNVGGTGSSTDLED--SKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYA 186
Db       396 VTTDSYGVATTYTSSSTADNAAKLNVQPTNN-TFQD------FDINYNGDT--------- 439

Qy       187 SRVMPSLNKVTTLFVAPQ----------CENGYTSGTMGFSSSNGDVAIDCSNIHIGITK 236
Db       440 --------KVMTVKYAGQTWTRNISDWIAKSGTTNFSLSMTASTGG-ATNLQQVQFG--- 487

Qy       237 GLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAG--------YRPFIDAYISATDVNQY 288
Db       488 ------------TFEYTESAVT---QVRYVDVTTGKDIIPPKTYSGNVDQVVTIDNQQSA 532
```

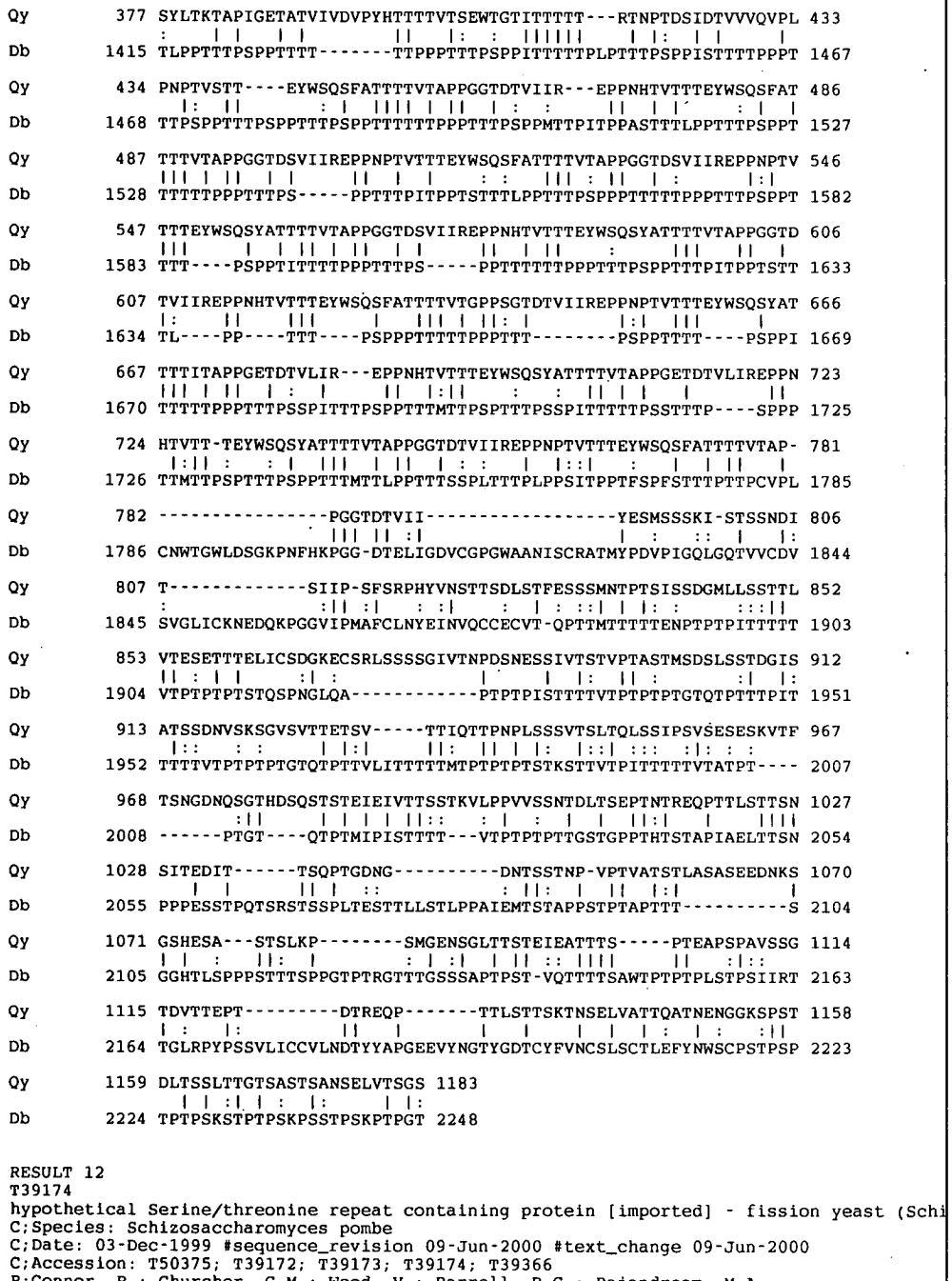


RESULT 10
T45025
mucin MUC5B, tracheobronchial [imported] - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T45025
R;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
J. Biol. Chem. 272, 3168-3178, 1997
A;Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alter
A;Reference number: Z22899; MUID:97166151; PMID:9013550
A;Accession: T45025
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3570 <DES>
A;Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
A;Experimental source: placenta
C;Genetics:
A;Gene: MUC5B

Query Match 9.7%; Score 632.5; DB 2; Length 3570;
Best Local Similarity 24.9%; Pred. No. 1.7e-19;
Matches 340; Conservative 155; Mismatches 509; Indels 363; Gaps 61;

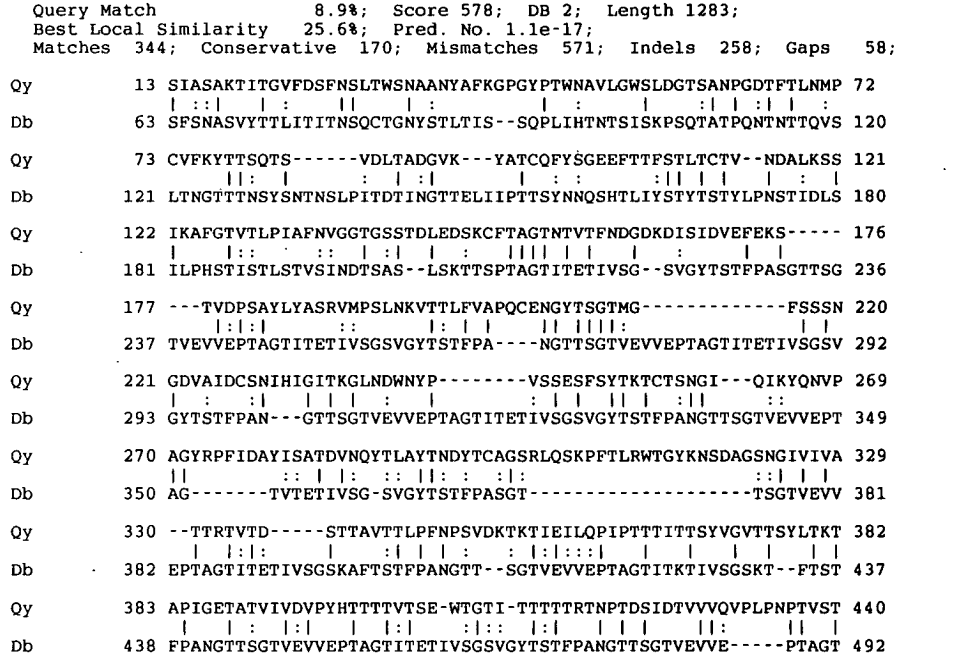


```
Db       2873 TATTTASTGSTATPSSTPGTAPPPKVLTSPATTPTATSSKATSSSSPRTAT-TLP----- 2926

Qy        571 SVIIREPPNHTVTT-TEYWSQSYATTTTV----TAPPGGTDTVIIREPPNHTVTTTEYWS 625

Db       2927 -VLTSTATKSTATSVTPIPSSTLGTTGTLPEQTTTPVATMSTIHPSSTPETTHTSTVLTT 2985

Qy        626 QSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREP 685

Db       2986 KATTRATSSTSTPSST--------PGTTWILTE-----LTTAATTTAGTGPTAT------ 3026

Qy        686 PNHTVTTTEYWSQSYATTT-TVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTA 744

Db       3027 PSSTPGTT--WILTELTTTATTTASTGSTAT--LSSTPGTTWILTE-----PSTTATVTA 3077

Qy        745 PPGGTDTVIIREPP--NPTVTTTEYWSQSFATTTTVTA----PPGGTDTVIIYESMSSSK 798

Db       3078 PPGSTATASSTQATAGTPHVSTT-------ATTPTVTSSKATPSSSPGTATALPALRSTA 3130

Qy        799 ISTSSNDITSIIPSFSRPHYVNSTTSDLS--TFESSSMNTPTSISSDGMLLSSTTLVTES 856

Db       3131 TTPTATSFTA-IPSSS----LGTTWTRLSQTTTPTATMSTATPSSTPETVHTSTVLTTTA 3185

Qy        857 ETTTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISATSS 916

Db       3186 TTT---------------GATGSVATPSSTPGTAHTTKVPT--------TTTTGFTATPS 3222

Qy        917 DNVSKSGVSVT----TETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGD 972

Db       3223 ---SSPGTALTPPVWISTTTTPTTTTPTTSGSTVTP----SSIPGT-------------- 3261

Qy        973 NQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITED 1032

Db       3262 ----THTAR-------VLTTTTT----TVATGSMATPSSSTQTSGTPPSLTTTATTITAT 3306

Qy       1033 ITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTT 1092

Db       3307 GSTTNPS-----STPGTTPIPPVLTSMATTPAATSSK------ATSSSSPRTATTLPVLT 3355

Qy       1093 STEIEATTTSPTEAPSPAVSSGTDV---TTEPTDTREQPTTLSTTSKTNSELVATTQATN 1149

Db       3356 STATKSTATSFTPIPSSTLWTTWTVPAQTTTPMSTMSTIHTSSTPETTHTSTVLTTTATM 3415

Qy       1150 ENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTS-------- 1201

Db       3416 T---RATNSTATPSSTLGT---TRILTELTTTATTTAATGSTATLSSTPGTTWILTEPST 3469

Qy       1202 ------VTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSI 1242

Db       3470 IATVMVPTGSTATTSSTLGTAHTPKVVTAMATMPTATASTVPSSSTV 3516


RESULT 11
A43932
mucin 2 precursor, intestinal - human (fragments)
N;Alternate names: mucin SMUC-41
C;Species: Homo sapiens (man)
C;Date: 10-Mar-1993 #sequence_revision 12-Apr-1996 #text_change 05-Nov-1999
C;Accession: A49963; A45106; B45106; A43932; B33532; A61257; PQ0328; PQ0329
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.
J. Biol. Chem. 269, 2440-2446, 1994
A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the
A;Reference number: A49963; MUID:94132002; PMID:8300571
A;Accession: A49963
A;Molecule type: mRNA
A;Residues: 1-639 <GU1>
A;Cross-references: GB:L21998
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Rothe, E.M.; Lagace, R.E.; Kim, Y.S.
J. Biol. Chem. 267, 21375-21383, 1992
A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr
A;Reference number: A45106; MUID:93016075; PMID:1400449
A;Accession: A45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 626-1895 <GU2>
A;Cross-references: GB:M94131; NID:g186395; PIDN:AAA59163.1; PID:g186396
A;Note: sequence extracted from NCBI backbone (NCBIP:116706)
A;Accession: B45106
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 2037-3020 <GU3>
A;Cross-references: GB:M94132; NID:g186397; PIDN:AAA59164.1; PID:g186398
A;Experimental source: colon
A;Note: sequence extracted from NCBI backbone (NCBIP:116698)
R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,
J. Clin. Invest. 88, 1005-1013, 1991
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym
A;Reference number: A43932; MUID:91358717; PMID:1885763
A;Accession: A43932
A;Molecule type: DNA
A;Residues: 1343-1350,'L',1352-1411,'S',1413-1448,'P',1450-1503,'T',1505-1915 <TOR>
A;Cross-references: GB:M74027; NID:g188863; PIDN:AAA59875.1; PID:g188864
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBIP:55750)
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.
J. Biol. Chem. 264, 6480-6487, 1989
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi
A;Reference number: A33532; MUID:89197956; PMID:2703501
A;Accession: B33532
A;Molecule type: mRNA
A;Residues: 1916-2193 <GU4>
A;Cross-references: GB:M22405; NID:g188873; PIDN:AAA36334.1; PID:g188874
A;Experimental source: intestine
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.
J. Clin. Invest. 87, 77-82, 1991
A;Title: Human bronchus and intestine express the same mucin gene.
A;Reference number: A61257; MUID:91086481; PMID:1985113
A;Accession: A61257
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 'T',1925-1948,'TTS',1952-1954 <JAN>
A;Experimental source: bronchus
R;Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstn
Biochem. Biophys. Res. Commun. 183, 821-828, 1992
A;Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the
A;Reference number: PQ0328; MUID:92198477; PMID:1550588
A;Accession: PQ0328
A;Molecule type: mRNA
A;Residues: 2328-2468 <XUG>
A;Cross-references: GB:M86523
A;Experimental source: small intestine
A;Accession: PQ0329
A;Molecule type: protein
A;Residues: 2328-2342,'K',2344-2354 <XUG1>
C;Genetics:
A;Gene: GDB:MUC2
A;Cross-references: GDB:120203; OMIM:158370
A;Map position: 11p15.5-11p15.5
C;Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; v
C;Keywords: glycoprotein; intestine; tandem repeat
F;2766-2834/Domain: von Willebrand factor type C repeat homology <VWC>

  Query Match             9.2%;  Score 599;  DB 2;  Length 3020;
  Best Local Similarity   25.8%;  Pred. No. 3.9e-18;
  Matches  285;  Conservative  120;  Mismatches  438;  Indels  262;  Gaps   44;

Qy        249 SFSYTKTCTSNGIQIKYQNV------PAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSR 302

Db       1236 AFCYWEICGPNGTVEKHFNICSITTRPSTLTTFTTITLPTTPTS-FTTTTTTTTPTSSTV 1294

Qy        303 LQSKP-FTLRWTGYKNSDAGSNGI----------VIVA----TTRTVTDSTTAVT----- 342

Db       1295 LSTTPKLCCLWSDWINEDHPSSGSDDGDREPFDGVCGAPEDIECRSVKDPHLSLEQHGQK 1354

Qy        343 -----------------TLPFNPSVDKTKTIE--------ILQPIPTTTITTSYVGVTT 376

Db       1355 VQCDVSVGFICKNEDQFGNGPFGLCYDYKIRVNCCWPMDKCITTPSPPTTTPSPPPTTTT 1414
```



RESULT 12
T39174
hypothetical Serine/threonine repeat containing protein [imported] - fission yeast (Schi
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C;Accession: T50375; T39172; T39173; T39174; T39366
R;Connor, R.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A;Reference number: Z21832
A;Accession: T50375
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1283 <CON>
A;Cross-references: EMBL:AL021815; PIDN:CAB61533.1; GSPDB:GN00067; SPDB:SPBC8E4.70c
A;Experimental source: strain 972h-; cosmid c8E4
A;Accession: T39172
A;Molecule type: DNA
A;Residues: 785-1283 <CO2>
A;Cross-references: EMBL:AL021815; PIDN:CAA17000.1
A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL
A;Accession: T39173
A;Molecule type: DNA
A;Residues: 'ME',179,'PLV',183,'W','KL',556-761,'HRGSS' <CO3>
A;Cross-references: EMBL:AL021815; PIDN:CAA17001.1
A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL
A;Accession: T39174
A;Molecule type: DNA
A;Residues: 1-555,'S' <CO4>
A;Cross-references: EMBL:AL021815; PIDN:CAA17002.1
A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Oliver, K.; Harris, D.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z21848
A;Accession: T39366
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-896,'G',898-904,'I',906-1283 <WOO>
A;Cross-references: EMBL:AL035675; PIDN:CAB38695.1; GSPDB:GN00067; SPDB:SPBC1289.15
A;Experimental source: strain 972h-; cosmid c1289
C;Genetics:
A;Gene: SPDB:SPBC1289.15; SPDB:SPAC8E4.07c
A;Map position: 2
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida

Query Match 8.9%; Score 578; DB 2; Length 1283;
Best Local Similarity 25.6%; Pred. No. 1.1e-17;
Matches 344; Conservative 170; Mismatches 571; Indels 258; Gaps 58;



```
Qy       441 TEYWSQSFATTTTVTAPPGGT--DTVIIREPPNHTVT-TTEYWSQSFATTTTVTAPPGGT 497
Db       493 ITETIVSGSKTFTSTFPASGTTSGTVEVVEPTAGTITETIVSGSKAFTSTFPANGTTSGT 552

Qy       498 DSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDS--VIIREPPNPTVTTTEYWSQS 555
Db       553 VEVV--EPTAGTITET---IVSGSVGYTSTFPASGTTSGTVEVVEPTAGTVTET---IVS 604

Qy       556 YATTTTVTAPPGGTDS--VIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREP 613
Db       605 GSVGYTSTFPASGTTSGTVEVVEPTAGTVTET-IVSGSVGYTSTFPASGTTSGTVEVVEP 663

Qy       614 PNHTVT-TTEYWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITA 672
Db       664 TAGTITETIVSGSKAFTSTFPANGTTSGTVEVV--EPTAGTITET-IVSGSKTFTSTFPA 720

Qy       673 PPGETDTVLIREPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYW 732
Db       721 NGTTSGTVEVVEPTAGTITET-IVSGSVGYTSTFPASGTTSGTVEVVEPTAGTVTETVI- 778

Qy       733 SQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYE 792
Db       779 SGSVSFMSTITAHDTSSGAVIVVEPTAGTVTET-IVSGSIPFTSTIPAQGTTSGTVEVVE 837

Qy       793 SMS---SSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSS 849
Db       838 PTAGTVTETIISGSVGYTSTFPA-------QGTTSGTVEVVEPTAGTVTETIISGSVGYT 890

Qy       850 TTLVTESET--TTELICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSS 907
Db       891 STFPAQSTTSGTVEMV----------APTAGTVTETIVSGSIPFTSTIPAQGTTSGTVEV 940

Qy       908 TDGISATSSDNVSKSGVSVTTE-----TSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESE 962
Db       941 VEPTAGTVTETIISGSVGYTSTFPAQGTTSGTVEIVA-PTAGTVTETIVSGSIP------ 993

Qy       963 SKVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTL 1022
Db       994 ----FTSTIPAQG------TTSGTVEVVEPTAGTVTETIISGSVGYTS---TFPAQGTTS 1040

Qy      1023 ST------TSNSITEDITTSQPTGDNGDNTSSTNPVPTVATST------LASASEEDNKS 1070
Db      1041 GTVEVVEPTAGTVTETIVSG--------SIPFTSTIPAQGTTSGTVEIVVPTAGTVTSTT 1092

Qy      1071 GSHESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSS-------------GTD 1116
Db      1093 GSGTSWFTTTVPATGTRSG-------SVIVVSPT---APACSSTPFAECPNLNFNFHSKN 1142

Qy      1117 VTTEPTDTREQPTTLSTTSKTNSELVATTQATNENGGKSPSTDLTSSLTTGTSASTSANS 1176
Db      1143 VNFPPADIR----VISVTPKGNNMYDATVQ----------------FTTSSTMSKKSLS 1181

Qy      1177 ELVTSGSVTGGAVASASNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTFIAS-- 1234
Db      1182 ELKILG------LSQTYLLYSYNSKVDN----ISNPGSWT--STVTVQGTSTGSYICMPH 1229

Qy      1235 ---TYD----GSGSIIQHSTWLY 1250
Db      1230 FQIQYDWCSAGVTDMSECNTWSY 1252
```

RESULT 13
S57180
probable membrane protein YJR151c - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein J2223; serine/threonine-rich protein YJR151c
C;Species: Saccharomyces cerevisiae
C;Date: 23-Aug-1995 #sequence_revision 08-Sep-1995 #text_change 19-Apr-2002
C;Accession: S57180
R;Scarcez, T.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S57169
A;Accession: S57180
A;Molecule type: DNA
A;Residues: 1-1161 <SCA>
A;Cross-references: EMBL:Z49651; NID:g1015902; PID:g1015903; GSPDB:GN00010; MIPS:YJR1
C;Genetics:
A;Gene: SGD:DAN4; MIPS:YJR151c
A;Cross-references: SGD:S0003912
A;Map position: 10R
C;Keywords: transmembrane protein

Query Match 8.9%; Score 577; DB 2; Length 1161;
Best Local Similarity 26.3%; Pred. No. 1.1e-17;
Matches 293; Conservative 131; Mismatches 427; Indels 264; Gaps 46;

```
Qy       276 IDAYISATDVNQYTLAY--------TNDY--TCAGSRLQSKPFTLRWTGYKNSDAGSNGI 325
Db        36 IELAVYVSDIRAHIFQYYSFRNHHKTETYPSEIAAAVFDYGDFTTRLTGISGDE------ 89

Qy       326 VIVATTRTVTDSTTAVTTLPFNPSVDKTKTIE-ILQPIPTTTITTSYVGVTTSYLTKTAP 384
Db        90 ----VTRMITGVPWYSTRL--KPAISSALSKDGIYTAIPTSTSTT-----TTKSSTSTTP 138

Qy       385 IGETATVIVDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYW 444
Db       139 ---------------TTTITS------TTSTTSTTPTTSTTST-------TPTTSTT--- 167

Qy       445 SQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIRE 504
Db       168 STTPTTSTTSTTPTTSTTSTTPTTSTTSTTPTTSTTSTTPTTSTTSTTPTT 227

Qy       505 PPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTA 564
Db       228 STTPTTSTTSTTSQTSTKSTTPTTSSTST---------TPTTSTTP------TTSTTSTA 272

Qy       565 PPGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW 624
Db       273 PTTSTTS---------TTSTTS--TISTAPTTSTTSSTFSTSSA----SASSVISTTATT 317

Qy       625 SQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIRE 684
Db       318 STTFASLTTPATSTASTDHT------TSSVSTTNAFTTSATTTTT-------SDTYISSS 364

Qy       685 PPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQ---------- 734
Db       365 SPSQVTSSAEPTTVSEVTSS--VEPTRSSQVTSSAEP----TTVSEFTSSVEPTRSSQVT 418

Qy       735 SYATTTTV---TAPPGGTDTVIIREPPNPTVTTTEYWSQ----------SFATTTTVTAP 781
Db       419 SSAEPTTVSEFTSSVEPTRSSQVTSSAEPT-TVSEFTSSVEPTRSSQVTSSAEPTTVSEF 477

Qy       782 PGGTDTVIIYESMSSSKISTSSNDITSIIP-------SFSRPHYVNSTTSDLSTFESSSM 834
Db       478 TSSVEPTRSSQVTSSAEPTTVSEFTSSVEPIRSSQVTSSAEPTTVSEVTSSVEPIRSSQV 537

Qy       835 NTPTSISSDGMLLSSTTLVTE----SETTT--------------ELICSDGKECSRLSSS 876
Db       538 TTTEPVSSFGSTFSEITSSAEPLSFSKATTSAESISSNQITISSELIVS-----SVITSS 592

Qy       877 SGIVTNPDSNESSIVTSTV-PTA---STMSDSLSSTDGISATS---------------- 915
Db       593 SEIPSSIEVLTSSGISSSVEPTSLVGPSSDESISSTESLSATSTFTSAVVSSSKAADFFT 652

Qy       916 -SDNVSKSGVSVTTETSVTTIQTTPN-PLSSSVTSLTQ----------LSSIPSVSESES 963
Db       653 RSTVSAKSDVSGNSSTQSTTFFATPSTPLAVSSTVVTSSTDSVSPNIPFSEISSSPESST 712

Qy       964 KVTFTSNGDNQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNT-----REQ 1018
Db       713 AITSTSTSFIAERTSSLYLSSSNMSSFTLSTFTVSQSIVSS---FSMEPTSSVASFASSS 769

Qy      1019 PTTLSTTSNSITEDITTSQPTGDNGDNTSSTNPVPTV-----ATSTLASASEEDNKSGSH 1073
Db       770 PLLVSSRSN--CSDARSS--------NTISSGLFSTIENVRNATSTFTNLSTD------- 812

Qy      1074 ESASTSLKPSMGENSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTT--EPTDTREQPTTL 1131
```

```
Db        813 EIVITSCKSSCTNEDSVLTKTQVSTVETTITSC------SGGICTTLMSPVTTINAKANT 866

Qy       1132 STTSKTNSELVATTQATNENGGKS----PSTDLTSSLTTGTSASTSANSELVTS------ 1181

Db        867 LTTTETST--VETTITTCPGGVCSTLTVPVTTITSEATTTATISCEDNEEDITSTETELL 924

Qy       1182 ------GSVTGGAVASASN-----DQSHSTSVTNSNSIVSNTPQT-------TLSQQVTS 1223

Db        925 TLETTITSCSGGICTTLMSPVTTINAKANTLTTTETSTVETTITTCSGGVCSTLTVPVTT 984

Qy       1224 -SSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLS 1257

Db        985 ITSEATTTATISCEDNEEDVASTKTELLTMETTIT 1019


RESULT 14
T32271
hypothetical protein ZC178.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T32271
R;Murray, J.; Wohldmann, P.; Beck, C.
submitted to the EMBL Data Library, September 1997
A;Description: The sequence of C. elegans cosmid ZC178.
A;Reference number: Z21143
A;Accession: T32271
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1459 <MUR>
A;Cross-references: EMBL:AF024496; PIDN:AAB70340.1; GSPDB:GN00023; CESP:ZC178.2
A;Experimental source: strain Bristol N2; clone ZC178
C;Genetics:
A;Gene: CESP:ZC178.2
A;Map position: 5
A;Introns: 1099/2

  Query Match             8.5%;  Score 553.5;  DB 2;  Length 1459;
  Best Local Similarity   24.5%;  Pred. No. 1.5e-16;
  Matches  333;  Conservative  172;  Mismatches  593;  Indels  261;  Gaps   55;

Qy         57 DGTSANPGDTFT----LNMPCVFKYTTSQTSVDLTADGVKYAT---------------- 95

Db        158 DSTTAGPSSTATNSASSETPC-----NSETQ---TSDGTSTMTVPNDSTTAGPSSTATNS 209

Qy         96 ------CQFYSGEEFTTFSTLTCTV-NDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDS 148

Db        210 ASSETPC---NSETQTSDGTSTMTVPNDS---------TTAGPSSTATNSASSETPCNSE 257

Qy        149 KCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENG 208

Db        258 TQTSDGTSTMT-----------VPNDSTTAGPS-----STATNSASSETPCNSETQTSDG 301

Qy        209 YTSGTMGFSSSNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNG---IQIKY 265

Db        302 TSTMTVPNDSTTAGPSSTATN--------------SASSETPCNSETQTSDGTSTMTVPN 347

Qy        266 QNVPAGYRPFIDAYISA-----------TDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTG 314

Db        348 DSTTAG--PSSTATNSASSETPCNSETQTSDGTSTMTVPNDSTTAGPSSTATNSASSETP 405

Qy        315 YKNSDAGSNGIVIVATTRTV-TDSTTA-----VTT-----LPFNPSV---DKTKTIEI-- 358

Db        406 CNSETQTSDG----TSTMTVPNDSTTAGPSSTVTNSASSETPCNSETQTSDGTSTMTVSS 461

Qy        359 --LQPIPTTTITTSYVGVT-------TSYLTKTAPI-------GETATVIVDVPYHT--- 399

Db        462 DSTTPGPSSTATNSASSETPCNSETQTSDGTFTMTVSSDSTTAGPSSTVTNSASSETPCN 521

Qy        400 TTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAP-- 457

Db        522 SETQTSDGTSTMTVSSDSTTPGPS-STATNSASSETPCNSET----QTSDGTSTMTVPND 576

Qy        458 -----PGGTDT--VIIREPPNHTVTTTEYWSQSFATTTTVTAPPGG--TDSVIIREPPNP 508

Db        577 STTAGPSSTATNSASSETPCNSETQTSDGTSTMTVPNDSTTAGPSSTVTNSASSETPCNS 636

Qy        509 TVTTTEYWSQSFATTTTVTAPPG--GTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPP 566

Db        637 ETQTSDGTSTMTVSSDSTTAGPSSTATNSASSETPCNSETQTSDGTSTMTVSSDSTTAGP 696

Qy        567 G--GTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDT-VIIREPP--NHTVTTT 621

Db        697 SSTATNSASSETPCNSETQTSDGTSTMTVSSDSTTAGPSSTATNSASSETPCNSETQTSD 756

Qy        622 EYWSQSFATTTTVTGPPSG-TDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTV 680

Db        757 GTFTMTVPNDSTTAGPSSTVTNSASSETPCNSETQTSDGTSTMTVSNDSTTAGPSSTAT- 815

Qy        681 LIREPPNHTVTTTEYWSQSYATTTTV-----TAPPGETDT--VLIREPPNHTVTTTEYWS 733

Db        816 --NSASSETPCNSETQTSDGTFTMTVPNDSTTAGPSSTATNSASSETPCNSETQTSDGTS 873

Qy        734 QSYATTTTVTAPPGGTDT--VIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGT------ 785

Db        874 TMTVPNDSTTAGPSSTATNSASSETPCNSETQTSDGTSTMTVPNDSTTAGPSSTATNSAS 933

Qy        786 -DTVIIYESMSSSKISTSS--NDITSIIPSFSRPHYVNSTT---SDLSTFESSSMNTPTS 839

Db        934 SETPCNSETQTSDGTSTMTVPNDSTTAGPSSTATNSASSETPCNSETQTSDGTSTMTVPN 993

Qy        840 ISSDGMLLSSTTLVTESET--TTELICSDGKEC-----SRLSSSSGIVTNPDSNES--S 889

Db        994 DSTTAGPSSTATNSASSETPCNSETQTSDGTSTMTVPNDSTTAGPSSTVTNSASSETPCN 1053

Qy        890 IVTSTVPTASTMS-DSLSSTDGISATSSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTS 948

Db       1054 SETQTSDGTSTMTVSSDSTTPGPSSTATNSASCETPFVYSHKFCKYDSTTAGPSSTATNS 1113

Qy        949 LTQLSSIPSVSESE--------------------SKVTFTSNGDNQSGTHDSQSTSTEIE 988

Db       1114 AS--SETPCNSETQTSDGTSTMTVPNDSTTAGPSSTVTNSASSETPCNSETQTSDGTFTM 1171

Qy        989 IVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSITEDITTSQPTGDNGDNTSS 1048

Db       1172 TVSSDSTTAGPSSTVTNSASSETPCNSETQ-TSDGTFTMTVSSDSTTAGPSSTVTNSASS 1230

Qy       1049 TNP---------------VPTVATSTLASASEEDNKSGSHESASTSLKPSMGENSGLTTS 1093

Db       1231 ETPCNSETQTSDGTSTMTVPN--DSTTAGPSSTVTNSASSETPCNSETQTSDGTSTMTVS 1288

Qy       1094 TEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSELVATTQATNENGG 1153

Db       1289 S--DSTTAGPSSTVTNSASSETPCNSE-TQTSDGTSTMTVSSDSTTAGQSST-ATNSASS 1344

Qy       1154 KSP---------------STDLT-----SSLTTGTSASTSANSELVTSGSVTGGAVASA 1192

Db       1345 ETPCNSETQTSDGTSTMTVSSDSTTAGPSSTVTNSASSETPCNSETQTSDGTSTMTVSSD 1404

Qy       1193 SNDQSHSTSVTNSNSIVSNTPQTTLSQQVTSSSPSTNTF 1231

Db       1405 STTAGQSSTATNSAS--SETPCNSETQ--TSDKTSTMTF 1439


RESULT 15
S25345
probable membrane protein YCR089w - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YCR1102
C;Species: Saccharomyces cerevisiae
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000
C;Accession: S25345; S19504
R;Wilson, C.; Grisanti, P.; Frontali, L.
Yeast 8, 569-575, 1992
A;Title: The complete sequence of a 6146 bp fragment of Saccharomyces cerevisiae chro
A;Reference number: S25345; MUID:92397594; PMID:1523889
A;Accession: S25345
```

```
A;Molecule type: DNA
A;Residues: 1-1609 <WIL>
A;Cross-references: GB:X59720; EMBL:S43845; NID:g1907116; PIDN:CAA42254.1; PID:g1907227
R;Frontali, L.; Grisanti, P.
submitted to the Protein Sequence Database, March 1992
A;Reference number: S19504
A;Accession: S19504
A;Molecule type: DNA
A;Residues: 1-1609 <FRO>
A;Cross-references: EMBL:X59720; NID:g1907116; PID:e264634; PID:g1907227; MIPS:YCR089w
C;Genetics:
A;Gene: SGD:FIG2
A;Cross-references: SGD:S0000685; MIPS:YCR089w
A;Map position: 3R
C;Keywords: transmembrane protein
F;4-20/Domain: transmembrane #status predicted <TM1>
F;1592-1609/Domain: transmembrane #status predicted <TM2>

  Query Match             8.4%;  Score 547.5;  DB 2;  Length 1609;
  Best Local Similarity   22.5%;  Pred. No. 3e-16;
  Matches  332;  Conservative  232;  Mismatches  503;  Indels  411;  Gaps   66;

Qy         55 SLDGTSANPGDTFT-LNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGEEFTTFSTLTCT 113
              ||  || ||  ::|  |     | :   :::| |:  :  :|    | | | |  |:| :
Db        165 SLTFTSVNPSQSWTSFNSE---KSSALSSTIDFTSSEISGSTSP-KSLESFDTTGTITSS 220

Qy        114 VNDA-------------------------LKSSIKAF----GTVTLPIAFNVGGTGSST 143
               : :                       | |:|:|      : |:|   :   :   |
Db        221 YSPSPSSKNSNQTSLLSPLEPLSSSSGDLILSSTIQATTNDQTSKTIPTLVDATSSLPPT 280

Qy        144 DLEDSKCFTAGTNTVTFN--------DGDKDISIDVEFEKSTVDPSAYL----YASRVMP 191
                  |   |:|:::::: |         |: |:        :::||| :     |:|  :
Db        281 LRSSSMAPTSGSDSISHNFTSPPSKTSGNYDV-----LTSNSIDPSLFTTTSEYSSTQLS 335

Qy        192 SLNKVT--------------------TLFVAPQCENGYT-SGTMGFSSSNGDVAIDCSN 229
              |||: :                    |  : |    | | |  :| |::|     : :
Db        336 SLNRASKSETVNFTASIASTPFGTDSATSLIDPISSVGSTASSFVGISTANFSTQGNSNY 395

Qy        230 IHIGITKG---LNDW---NYPVS-------------------SESFSYTKTCTSNGIQI 263
              :     |     ||   : |:|                  | ::  | | | :|:
Db        396 VPESTASGSSQYQDWSSSSLPLSQTTWVVINTTNTQGSVTSTTSPAYVSTATKTVDGVIT 455

Qy        264 KY---------QNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLQSKPFTLRWTG 314
              :|        ::   |    | :   |:  :  ::  :|  | | |    : | :
Db        456 EYVTWCPLTQTKSQAIGVSSSISSVPQASSFSGSSILSSNSSTLAAS--NNVPESTASGS 513

Qy        315 YKNSDAGSNGIVIVATT-----RTVTDSTTAVTTLPFNPSVDKTKTI-----EILQPIPT 364
               :  |  |: : :  ||      | |  :   || |   |   |||:     | :   |
Db        514 SQYQDWSSSSLPLSQTTWVVINTTNTQGSVTSTTSPAYVST-ATKTVDGVITEYVTWCPL 572

Qy        365 TTITTSYVGVTTSYLTKT--------APIGETATVIVDVPYHTTTTVTSEWTGTITTTTT 416
              |   :  :|:::| :: |          :| :   : |  :  | |:    |  :| :|:
Db        573 TQTKSQAIGISSSTISATQTSKPSSILTLGISTLQLSDATFKGTETIN---THLMTESTS 629

Qy        417 RTNPT------DSIDTVVVQVPLPNPTVSTTEYWSQSFATTTTVTAPPGGTDTVIIREPP 470
               | ||      ||      :| | : ::|:   :| |  :| |:|
Db        630 ITEPTYFSGTSDSFYLCTSEVNLAS-SLSSYPNFSSSEGSTATIT-------------- 673

Qy        471 NHTVT---TTEYWSQSFATTTTVTA-----------PPGGTDSVIIREPPNPTVTTTEY 515
              | |||   |::| | | :  |  :                |::: :    : | :  :|
Db        674 NSTVTFGSTSKYPSTSVSNPTEASQHVSSSVNSLTDFTSNSTETIAVISNIHKTSSNKDY 733

Qy        516 WSQSFATTTTVTAPPGGTDSVIIREPPNPTVTT------TEY--W--SQSYATTTTVTAP 565
                : ||      |  ::::     ||||      |||  |  : | | ||:::
Db        734 ------SLTTTQLKTSGKQTLVL-----STVTTTVNGAATEYTTWCPASSIAYTTSISY- 781

Qy        566 PGGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTV---TTTE 622
                          | |||  | |  | | :|:    : |: :    : ||   | :|
Db        782 ------------KTLVLTTEVCSHSECTPTVITSVTATSSTIPLLSTSSSTVLSSTVSE 828

Qy        623 YWSQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLI 682
                    |:  |:    | |           | |:|:       : |: ||   |: |
Db        829 GAKNPAASEVTINTQVSATS--------EATSTSTQ------VSATSATATASESSTTSQ 874

Qy        683 REPPNHTVTTTEYWSQSYATTTTVTAPPGETD----TVLIREPPNHTVTTTEYWSQSYAT 738
                 : |::|    :|:: || ::  |   |:    ||: |:     : :||  | |
Db        875 VSTASETIST--LGTQNFTTTGSLLFPALSTEMINTTVVSRKT---LIISTEVCSHSKCV 929

Qy        739 TTTVT-------APPGG-------TDTVIIREPPNPTVT-TTEYWSQSFATTTTVTAPPG 783
               | :|       |  |       |: | :    : ||| :||  | |  | | :|:
Db        930 PTVITEVVTSKGTPSNGHSSQTLQTEAVEVTLSSHQTVTMSTEVCSNSICTPTVITSVQ- 988

Qy        784 GTDTVIIYESMSSSKISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSD 843
                  |   | : |:|  | :|   :|:             :|::|||  |: : | : :
Db        989 MRSTPFPYLTSSTSSSSLASTKKSSL----------EASSEMSTFSVSTQSLPLAFTCS 1037

Qy        844 GMLLSSTTLVTESETT--TELICSDGKECSRLSSSSGIVTN--PDSNES--------SIV 891
                    ||| |::   |  |  | |        |||: | ||  | |  |        |:
Db       1038 EK--RSTTSVSQWSNTVLTNTIMS--------SSSNVISTNEKPSSTTSPYNFSSGYSLP 1087

Qy        892 TSTVPTASTMSDSLSSTDGI------------------------------SATSSDNVS 920
               :|: |:  ::| : :: :||                            |:  | ::|
Db       1088 SSSTPSQYSLSTATTTINGIKTVYTTWCPLAEKSTVAASSQSSRSVDRFVSSSKPSSSLS 1147

Qy        921 KSGVSVTTETSVTTIQ------TTPNPLSSSVT--SLTQLSSIPSVSESESKVTFTSNGD 972
               :: :  |  |: |||       ||  ||:|  |  : || |       | :||  ||
Db       1148 QTSIQYTLSTATTTISGLKTVYTTWCPLTSKSTLGATTQTS-------STAKVRITS-AS 1199

Qy        973 NQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSE-PTNTREQPTTLS-----TTS 1026
               : : |  | ||||| |    ||:  |   | | |: | : || :    :||:     :||
Db       1200 SATSTSISLSTSTESE----SSSGYLSKGVCSGTECTQDVPTQSSSPASTLAYSPSVSTS 1255

Qy       1027 NSITEDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGE 1086
               :| :   ||:          ||:   || : :|:  |||   : |      ||||
Db       1256 SSSSFSTTTASTL-----TSTHTSVPLLPSSSSISASSPSSTS----LLSTSLPSPAFT 1305

Qy       1087 NSGLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSEL----- 1141
               :| | |:| : ::|   :  |   :|| : ::  |  :    :  |::| ::| |
Db       1306 SSTLPTATAVSSSTFIASSLP---LSSKSSLSLSPVSSSILMSQFSSSSSSSSSLASLPS 1362

Qy       1142 ------------------VATTQATNENGGKSPST--------DLTSSLT------TGTS 1169
                             :||   |:    :  ||  '       |:||: |      | |
Db       1363 LSISPTVDTVSVLQPTTSIATLTCTDSQCQQEVSTICNGSNCDDVTSTATTPPSTVTDTM 1422

Qy       1170 ASTSANSELVTSGSVTG-----------GAVASASND--QSHSTSVTNS-----NSIVS 1210
                | :  :  || |  |              ::: | :  |: :||  ||      |:::
Db       1423 TCTGSECQKTTSSSCDGYSCKVSETYKSSATISACSGEGCQASATSELNSQYVTMTSVIT 1482

Qy       1211 NTPQTTLSQQV--TSSSPSTNTFIASTY---DGSGSII 1243
                :  || | :|  | |: |  |    ||   | :| :|
Db       1483 PSAITTTSVEVHSTESTISITTVKPVTYTSSDTNGELI 1520

Search completed: June 11, 2003, 17:14:06
Job time : 42 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2004, 20:45:33 ; Search time 4785 Seconds
(without alignments)
10644.198 Million cell updates/sec

Title: US-09-715-876-7_COPY_52_1296
Perfect score: 1245
Sequence: 1 aagacaatcactggtgtttt.....cagtggtggtacaagttcca 1245

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1245	100.0	3786	8	YSAALS1	L25902 Candida alb
	2	1046.6	84.1	3813	8	AF068866	AF068866 Candida a
	3	1037	83.3	4723	8	AF025429	AF025429 Candida a
	4	1030.6	82.8	4044	8	AY227439	AY227439 Candida a
	5	1029	82.7	4152	8	AY227440	AY227440 Candida a
	6	949	76.2	3144	8	AY223551	AY223551 Candida a
	7	949	76.2	4383	8	AF051313	AF051313 Candida a
	8	947.4	76.1	3360	8	CAU87956	U87956 Candida alb
	9	945.8	76.0	3468	8	AY223552	AY223552 Candida a
	10	895	71.9	1404	8	CA1ALS1	AF024580 Candida a
	11	890.2	71.5	1404	8	CANALS2S1	AF024582 Candida a
	12	733.4	58.9	1407	8	CANALS4S1	AF024584 Candida a
	13	733	58.9	4569	8	AF272027	AF272027 Candida a
	14	731.8	58.8	1007	8	AF202529	AF202529 Candida d
	15	728.6	58.5	1407	8	CANALS42S1	AF024586 Candida a
	16	713.8	57.3	1404	8	AF229989S1	AF229989 Candida a
	17	713.8	57.3	5502	8	AY269422	AY269422 Candida a
	18	713.8	57.3	5565	8	AY269423	AY269423 Candida a
	19	574.6	46.2	4332	8	AF075293	AF075293 Candida a
	20	573	46.0	4101	8	AY225310	AY225310 Candida a
	21	507	40.7	995	8	AF202530	AF202530 Candida d
	22	463.6	37.2	1013	8	AF201685	AF201685 Candida d
	23	393.6	31.6	6897	8	AF201684	AF201684 Candida a
	24	392	31.5	1413	8	AY296650	AY296650 Candida a
	25	390.6	31.4	1299	8	AF075294	AF075294 Candida a
	26	362.2	29.1	1058	8	AF201686	AF201686 Candida t
	27	141.6	11.4	377	8	AF211866	AF211866 Candida t
	28	132	10.6	380	8	AF211865	AF211865 Candida t
	29	125.4	10.1	1400	8	AB002099	AB002099 Candida t
	30	112.2	9.0	157	8	AF035757	AF035757 Candida a
	31	105	8.4	170221	10	AC127421	AC127421 Mus muscu
	32	103.4	8.3	216959	10	AC116998	AC116998 Mus muscu
	33	90.2	7.2	255220	2	AC122672	AC122672 Rattus no
	34	88.2	7.1	170669	2	BX322580	BX322580 Danio rer
c	35	87.8	7.1	146907	2	AC141031	AC141031 Rattus no
c	36	87.8	7.1	158615	2	AC117835	AC117835 Rattus no
	37	86.8	7.0	203271	2	BX322566	BX322566 Danio rer
c	38	85.6	6.9	176322	2	AL935062	AL935062 Danio rer
	39	85	6.8	35793	5	AY016024	AY016024 Takifugu
	40	83.8	6.7	266451	2	AC098493	AC098493 Rattus no
c	41	82.4	6.6	100	6	A43858	A43858 Sequence 32
c	42	82.4	6.6	100	6	A43860	A43860 Sequence 34
c	43	82.4	6.6	100	6	I17435	I17435 Sequence 32
c	44	82.4	6.6	100	6	I17437	I17437 Sequence 34
c	45	82	6.6	158266	2	BX323881	BX323881 Danio rer

ALIGNMENTS

RESULT 1
YSAALS1
LOCUS YSAALS1 3786 bp DNA linear PLN 03-MAY-2000
DEFINITION Candida albicans agglutinin-like sequence (ALS1) gene, complete cds.
ACCESSION L25902
VERSION L25902.1 GI:704426
KEYWORDS .
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 3786)
AUTHORS Hoyer,L.L., Scherer,S., Shatzman,A.R. and Livi,G.P.
TITLE Candida albicans ALS1: domains related to a Saccharomyces

```
          cerevisiae sexual agglutinin separated by a repeating motif
  JOURNAL   Mol. Microbiol. 15 (1), 39-54 (1995)
  MEDLINE   95272392
   PUBMED   7752895
FEATURES             Location/Qualifiers
     source          1. .3786
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /strain="B792"
                     /db_xref="taxon:5476"
     gene            1. .3783
                     /gene="ALS1"
     CDS             1. .3783
                     /gene="ALS1"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAC41649.2"
                     /db_xref="GI:7687905"
                     /translation="MLQQFTLLFLYLSIASAKTITGVFDSFNSLTWSNAANYAFKGPG
                     YPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKYTTSQTSVDLTADGVKYATCQFYSGE
                     EFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGTNTVTF
                     NDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGTMGFSS
                     SNGDVAIDCSNIHIGITKGLNDWNYPVSSESFSYTKTCTSNGIQIKYQNVPAGYRPFI
                     DAYISATDVNQYTLAYTNDYTCAGSRSQSKPFTLRWTGYKNSDAGSNGIVIVATTRTV
                     TDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPSPNPTVSTTEYWSQSFAT
                     TTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNP
                     TVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPP
                     GGTDSVIIREPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNHTVTTTEYW
                     SQSFATTTTVTGPPSGTDTVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLI
                     REPPNHTVTTTEYWSQSYATTTTVTAPPGETDTVLIREPPNHTVTTTEYWSQSYATTT
                     TVTAPPGGTDTVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDTVIIYESMSSSK
                     ISTSSNDITSIIPSFSRPHYVNSTTSDLSTFESSSMNTPTSISSDGMLLSSTTLVTES
                     ETTTESICSDGKECSRLSSSSGIVTNPDSNESSIVTSTVPTASTMSDSLSSTDGISAT
                     SSDNVSKSGVSVTTETSVTTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVTFTSNGD
                     NQSGTHDSQSTSTEIEIVTTSSTKVLPPVVSSNTDLTSEPTNTREQPTTLSTTSNSIT
                     EDITTSQPTGDNGDNTSSTNPVPTVATSTLASASEEDNKSGSHESASTSLKPSMGENS
                     GLTTSTEIEATTTSPTEAPSPAVSSGTDVTTEPTDTREQPTTLSTTSKTNSESVATTQ
                     ATNENGGKSPSTDLTSSLTTGTSASTSANSELVTSGSVTGGAVASASNDQSHSTSVTN
                     SNSIVSNTPQTTLSQQVTSSSPSTNTFIASTYDGSGSIIQHSTWLYGLITLLSLFI"
     repeat_region   1299. .2370
                     /rpt_unit=1299. .1407
BASE COUNT     1170 a    899 c    592 g   1125 t
ORIGIN

  Query Match           100.0%;  Score 1245;  DB 8;  Length 3786;
  Best Local Similarity  100.0%;  Pred. No. 4.6e-219;
  Matches 1245;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAGACAATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         52 AAGACAATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAAT 111

Qy         61 TATGCTTTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        112 TATGCTTTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGAT 171

Qy        121 GGTACCAGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        172 GGTACCAGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATAT 231

Qy        181 ACTACTTCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        232 ACTACTTCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAA 291

Qy        241 TTTTATTCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        292 TTTTATTCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCT 351

Qy        301 TTGAAATCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 TTGAAATCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGT 411

Qy        361 GGAACAGGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        412 GGAACAGGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACA 471

Qy        421 GTCACATTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        472 GTCACATTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACC 531

Qy        481 GTTGATCCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        532 GTTGATCCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACA 591

Qy        541 ACTCTTTTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        592 ACTCTTTTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCC 651

Qy        601 AGTAGTAACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGA 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        652 AGTAGTAACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGA 711

Qy        661 TTAAATGATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        712 TTAAATGATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACA 771

Qy        721 TCTAATGGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        772 TCTAATGGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGAT 831

Qy        781 GCTTATATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        832 GCTTATATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACT 891

Qy        841 TGTGCTGGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        892 TGTGCTGGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAAT 951

Qy        901 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        952 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 1011

Qy        961 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1012 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1071

Qy       1021 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1072 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1131

Qy       1081 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1132 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1191

Qy       1141 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1192 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGT 1251

Qy       1201 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1245
              |||||||||||||||||||||||||||||||||||||||||||||
Db       1252 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1296


RESULT 2
AF068866
LOCUS       AF068866               3813 bp    DNA     linear   PLN 29-JAN-2001
DEFINITION  Candida albicans agglutinin-like protein (ALS5) gene, ALS5-1
            allele, complete cds.
ACCESSION   AF068866
VERSION     AF068866.1  GI:4903268
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
```

```
          Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 3813)
  AUTHORS   Hoyer,L.L. and Hecht,J.E.
  TITLE     The ALS5 gene of Candida albicans and analysis of the Als5p
            N-terminal domain
  JOURNAL   Yeast 18 (1), 49-60 (2001)
  MEDLINE   21064501
   PUBMED   11124701
REFERENCE   2  (bases 1 to 3813)
  AUTHORS   Hoyer,L.L., Ho,M. and Hecht,J.E.
  TITLE     The ALS5, ALS6 and ALS7 genes of Candida albicans
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 3813)
  AUTHORS   Hoyer,L.L., Ho,M. and Hecht,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) Veterinary Pathobiology, University of
            Illinois at Urbana-Champaign, 2001 S. Lincoln Avenue, Urbana, IL
            61802, USA
FEATURES             Location/Qualifiers
     source          1. .3813
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="3"
                     /note="SfiI fragment P"
     gene            1. .3813
                     /gene="ALS5"
                     /allele="ALS5-1"
     CDS             1. .3813
                     /gene="ALS5"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAD32849.1"
                     /db_xref="GI:4903269"
                     /translation="MIQQFTLLFLYLSFATAKAITGIFNSIDSLTWSNAGNYAFKGPG
                     YPTWNAVLGWSLDGTSANPGDTFILNMPCVFKFTASQKSVDLTADGVKYATCQFYSGE
                     EFTTFSTLTCTVNDALKSSIKAFGTVTLPIAFNVGGTGSSTDLEDSKCFTAGINTVTF
                     NDGSKKLSIAVNFEKSTVDRSGYLTTSRFMPSLNKIATLYVAPQCENGYTSGTMGFST
                     SYGDVAIDCSNVHIGISKGVNDWNHPVTSESFSYTKSCSSFGISITYQNVPAGYRPFI
                     DAYISPSDNNQYQLSYKNDYTCVDDYWQHAPFTLKWTGYKNSDAGSNGIVIVATTRTV
                     TDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATLI
                     VDVPYHTTTTVTSKWTGTITMTTTRTNPTDSIDTVVVQVPLPNPTTTTTQFWSESFTS
                     TTTITNKPEGTDSVIVKEPHNPTVTTTEFWSESYATTETITTGPLGTDSIVIHDPLEE
                     SSSTTAIESSDSNISSSAQESSSSVEQSSSIVGLSSSSDIPLSSDMPSSSSTGLTSSE
                     SSTVSSYDSDSSSSSSELSTFSSSESYSSSISDTTNFWDSSSSDLESTSITWSSSIDAQ
                     SSQSVQSVSNSISTSQETTSSSGEESNTSVTDILVSSDASSILNSDISSYYPSSTISL
                     SDDFPHTIAGEPDSRSSSSIASTVEISSDLVSLTSDPTSSFDSSSSLNSDSSSSPFSD
                     ESDISASSSFSTLVAPSFSLSSSSSLSLIYPHYVNSTTYHASESESSSVASPSVASES
                     ANDDTHTLSESTDTTSIIGTDSSTVTFCRRDNGDGCIVTGITSSSIDSEQTSDVTTTS
                     SFVASSTPTSAEQSITDNPNIDSSQTSASSSTKSSVSVSDTVVNSILLSETSTLSSDD
                     GTSSDTSISSTTNSDTGNINAGSSHKSTASIKESSIQKTGVTLSSSYLSTKLSSTSDI
                     TIELITTELTTIEDNEPNTFTSTPSSHSEIFSSDNSVLSKQVDRESTIKTSPTTDVTT
                     VSSLSVHSTEASTATLGENSFSNVASTPLNTATSLRSTSSSSNHATESSGTVKSEASV
                     EAIPSPPTSTDNRLSYSTEEAEGITYANSGSTNNLITESQVAAPTDSTSVLIENLVVT
                     STFDDNSSAAVDQPSKTKSIEESIMNPDSTNETNNGFIATLSQAQVPSSSIHSELIST
                     TTAKTTDASMNGDSAASNSQPTTLIQQVATSSYNQPLITTYAGSSSATKHPSWLLKFI
                     SVALFFFL"
BASE COUNT     1172 a    819 c    641 g   1181 t
ORIGIN

  Query Match             84.1%;  Score 1046.6;  DB 8;  Length 3813;
  Best Local Similarity   90.0%;  Pred. No. 1.3e-182;
  Matches 1121;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy          1 AAGACAATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAAT 60
              ||| | |||||||||| ||||  ||||| || | ||||||||||||||||||||  |||
Db         52 AAGGCGATCACTGGTATTTTCAATAGTATTGACTCATTAACTTGGTCCAATGCTGGCAAT 111

Qy         61 TATGCTTTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGAT 120
              || |||||||||| |||||||||||||||||||||||||||| |||||||||| ||||||
Db        112 TACGCTTTCAAAGGACCAGGATACCCAACTTGGAATGCTGTGTTGGGTTGGTCATTAGAT 171

Qy        121 GGTACCAGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATAT 180
              |||||||||||||||||||| |||||||||| || || ||||||||||||||||||||
Db        172 GGTACCAGTGCCAATCCAGGAGATACATTCATATTAAACATGCCATGTGTGTTTAAATTC 231

Qy        181 ACTACTTCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAA 240
              ||| |||| |||| |||||||||||| ||||||||||||||||||||||||||||||||
Db        232 ACTGCTTCCCAAAAATCTGTTGATTTGACTGCCGATGGTGTTAAATATGCTACTTGTCAA 291

Qy        241 TTTTATTCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCT 300
              |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||
Db        292 TTTTATTCTGGTGAAGAGTTTACAACTTTTTCTACATTAACATGTACTGTGAACGACGCT 351

Qy        301 TTGAAATCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        352 TTGAAATCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGT 411

Qy        361 GGAACAGGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACA 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        412 GGAACAGGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTATCAATACG 471

Qy        421 GTCACATTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACC 480
              || |||||||||||||   |||| |  ||||||||| |||| | ||||||||||||||
Db        472 GTAACATTTAATGATGGCAGTAAAAAGCTCTCAATTGCTGTTAATTTTGAAAAGTCAACA 531

Qy        481 GTTGATCCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACA 540
              ||||||| |||||  ||||||  | |||||||| ||||||| ||||||||||  | | |
Db        532 GTTGATCGAAGTGGGTATTTGACTACTTCCAGATTTATGCCGAGTCTCAATAAAATTGCT 591

Qy        541 ACTCTTTTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCC 600
              ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        592 ACTCTTTATGTGGCACCACAATGTGAAAACGGTTACACATCTGGTACAATGGGATTCTCC 651

Qy        601 AGTAGTAACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGA 660
              | |||| | || || |||||||||| || |||||| | |||||||||||  ||||||||
Db        652 ACTAGTTATGGGGATGTTGCTATTGACTGTTCAAATGTACATATTGGTATTTCAAAAGGA 711

Qy        661 TTAAATGATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACA 720
              |||||||||||||| |||| ||| | |||||||||||||||||||||||  ||| ||
Db        712 GTAAATGATTGGAATCATCCAGTTACGTCTGAATCATTTAGTTACACTAAAAGCTGTTCA 771

Qy        721 TCTAATGGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGAT 780
              ||| ||| ||    || | |||||||||||| |||||  ||||||||||||||||||
Db        772 TCTTTTGGTATCTCTATCACATATCAAAATGTTCCTGCCGGTTATCGTCCATTTATTGAC 831

Qy        781 GCTTATATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACT 840
              |||||||||||  |||||  |||||| ||| ||  | ||||  ||||| ||||||||||
Db        832 GCTTATATTTCTCCCTCAGATAATAACCAGTATCAATTGTCGTATAAAAATGACTATACT 891

Qy        841 TGTGCTGGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAAT 900
              |||| ||    |  | |||| | ||||||||||||| |||||||||||  ||||||
Db        892 TGTGTTGATGATTATTGGCAACATGCACCTTTCACTTTAAAATGGACTGGATATAAGAAT 951

Qy        901 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        952 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 1011

Qy        961 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1012 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1071

Qy       1021 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1072 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1131

Qy       1081 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1140
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       1132 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTCTTATTGTTGATGTGCCATAT 1191

Qy       1141 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGT 1200
              |||||||||||||||||| ||| |||||||||||||||| || |  || ||||||||||
Db       1192 CATACTACCACAACTGTTACAAGTAAATGGACAGGAACAATTACGATGACTACAACTCGT 1251
```

```
Qy       1201 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1245
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1252 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1296
```

RESULT 3
AF025429
LOCUS AF025429 4723 bp DNA linear PLN 16-DEC-1997
DEFINITION Candida albicans agglutinin-like adhesin (ALA1) gene, complete cds.
ACCESSION AF025429
VERSION AF025429.1 GI:2522218
KEYWORDS .
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 4723)
AUTHORS Gaur,N.K. and Klotz,S.A.
TITLE Expression, cloning, and characterization of a Candida albicans gene, ALA1, that confers adherence properties upon Saccharomyces cerevisiae for extracellular matrix proteins
JOURNAL Infect. Immun. 65 (12), 5289-5294 (1997)
MEDLINE 98053977
PUBMED 9393828
REFERENCE 2 (bases 1 to 4723)
AUTHORS Gaur,N.K. and Klotz,S.A.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1997) Research Service, VA Medical Center, 4801 Linwood Blvd., Kansas City, MO 64128, USA
FEATURES Location/Qualifiers
source 1. .4723
/organism="Candida albicans"
/mol_type="genomic DNA"
/db_xref="taxon:5476"
gene 327. .4586
/gene="ALA1"
CDS 327. .4586
/gene="ALA1"
/function="cell adhesion protein"
/codon_start=1
/transl_table=12
/product="agglutinin-like adhesin"
/protein_id="AAB88883.1"
/db_xref="GI:2522219"
/translation="MIQQFTLLFLYLSFATAKAITGIFNSIDSLTWSNAGNYAFKGPG
YPTWNAVLGWSLDGTSANPGDTFILNMPCVFKFTASQKSVDLTADGVKYATCQFYSGE
EFTTFSSLKCTVNNNLRSSIKALGTVTLPIAFNVGGTGSSVDLEDSKCFTAGTNTVTF
NDGSKKLSIAVNFEKSTVDQSGYLTTSRFMPSLNKIATLYVAPQCENGYTSGTMGFST
SYGDVAIDCSNVHIGISKGVNDWNHPVTSESFSYTKSCSSFGISITYQNVPAGYRPFI
DAYISPSDNNQYQLSYKNDYTCVDDYWQHAPFTLKWTGYKNSDAGSNGIVIVATTRTV
TDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
VDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPSPNPTTTTTQFWSESFTS
TTTITNSLKGTDSVIVREPHNPTVTTTEFWSESFATTETITSKPEGTDSVIVREPHNP
TVTTTEFWSESYATTETITNGPEGTDSVIVREPHNPTVTTTKFWSESYATTETITNKP
EGTDSVIVKEPYNPTVTTTEFWSESYATTETITNGPEGTDSVIVREPHNPTVTTTEFW
SESYATTETITTGPLGTDSIVIHDPLEESSSTTAIESSDSNISSSAQESSSSVEQSSS
IVGLSSSSDIPLSSDMPSSSSTGLTSSESSTVSSYDSDSSSSIESSTLSSSDRCSSSI
SDTTSFWDSSSSDLESTSITWSSSIDAQSSHLVQSVSNSISTSQELSSSSSEESSTFA
TDALVSSDASSILSSDTSSYYPSSTISSSDDFPHTIAGESDSLSISFITSTVEISSDS
VSLTSDPASSFDSSSSLNSDSSSSPSSDQSDILTSSSFSTLVVPSFSLSSSSSLSLTY
PHYVNSTTYHASESESSSVASPSMASESANDDTYTLSESTDTTSSIGTDSSTVTFCRR
DNGDGCIVTGMPSSSIDSEQTSDVTTTSSFVASSTPTSAEQSITDNPNIDSSQTSASS
STKSSVSVSDTVVNSILLSETSTLSSDDSTSSDTSISSTTNSDTGNINAGSSHTSTAS
IKESSIQKTGVTLSSSYLSTKLSSTSDITIELITTELITTELTTIEDNEPNTFTSTPS
SHSEIFSSDNSVLSKQVDRESTIKTSPTTDVTTVSSLSVHSTEASTATLGENSFSNVA
STPSNIATSLRSTSSSSNHATESSGTVKSEASAEAIPSPPTSTDNRLSYSTEEAKGIT
YANSGSTNNLITESQVAAPTDSTSVLIENPVVTSTFDDNSSAAVDQPSKTKSIEESIM
NPDSTNETNNGFIATLSQAQVPSSSIHSELISTTTAKTTDASMNGDSAASNSQPTTLI
QQVATSSYNQPLITTYAGSSSATKHPSWLLKFISVALFFFL"
BASE COUNT 1441 a 1013 c 803 g 1466 t
ORIGIN

Query Match 83.3%; Score 1037; DB 8; Length 4723;

Best Local Similarity 89.6%; Pred. No. 7e-181;
Matches 1115; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

```
Qy          1 AAGACAATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAAT 60
              ||| | |||||||||| ||||  ||||| || | ||||||||||||||||||||  |||
Db        378 AAGGCGATCACTGGTATTTTCAATAGTATTGACTCATTAACTTGGTCCAATGCTGGCAAT 437

Qy         61 TATGCTTTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGAT 120
              || |||||||||| ||||||||||||||||||||||||||| |||||||||| ||||||
Db        438 TACGCTTTCAAAGGACCAGGATACCCAACTTGGAATGCTGTGTTGGGTTGGTCATTAGAT 497

Qy        121 GGTACCAGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATAT 180
              |||||||||||||||||||||||||||||| ||| || ||||||||||||||||||
Db        498 GGTACCAGTGCCAATCCAGGGGATACATTCATATTAAACATGCCATGTGTGTTTAAATTC 557

Qy        181 ACTACTTCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAA 240
              ||| |||| |||| |||||||||||| ||||||||||||||||||||||||||||||||
Db        558 ACTGCTTCCCAAAAATCTGTTGATTTGACTGCCGATGGTGTTAAATATGCTACTTGTCAA 617

Qy        241 TTTTATTCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCT 300
              |||||||||||||||| || |||||||||||| |||||| |||||| |||||| |   |
Db        618 TTTTATTCTGGTGAAGAGTTTACAACTTTTTCTTCATTAAAATGTACAGTGAACAATAAT 677

Qy        301 TTGAAATCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGT 360
              |||| |||||| |||||||| || ||||| ||||||||||||||||||||||||||||
Db        678 TTGAGATCATCTATTAAGGCTTTGGGTACGGTTACTTTACCAATTGCATTCAATGTTGGT 737

Qy        361 GGAACAGGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACA 420
              |||||||| ||||||  |||||||||||||||||||||||||||||||||||||||||||
Db        738 GGAACAGGGTCATCAGTTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACG 797

Qy        421 GTCACATTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACC 480
              || ||||||||||||||    |||| |  |||||||| |||| | ||||||||||||||
Db        798 GTAACATTTAATGATGGCAGTAAAAAGCTCTCAATTGCTGTTAATTTTGAAAAGTCAACA 857

Qy        481 GTTGATCCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACA 540
              ||||||| |||||  ||||||  | |||||||| |||||| ||||||||||  |  |
Db        858 GTTGATCAAAGTGGGTATTTGACTACTTCCAGATTTATGCCGAGTCTCAATAAAATTGCT 917

Qy        541 ACTCTTTTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCC 600
              ||||||| |||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        918 ACTCTTTATGTGGCACCACAATGTGAAAACGGTTACACATCTGGTACAATGGGATTCTCC 977

Qy        601 AGTAGTAACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGA 660
              | |||| | || || |||||||||| || |||||| | |||||||||| ||||||||
Db        978 ACTAGTTATGGGGATGTTGCTATTGACTGTTCAAATGTACATATTGGTATTTCAAAAGGA 1037

Qy        661 TTAAATGATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACA 720
              |||||||||||||| |||| ||| | |||||||||||||||||||||| |||| ||
Db       1038 GTAAATGATTGGAATCATCCAGTTACGTCTGAATCATTTAGTTACACTAAAAGTTGTTCA 1097

Qy        721 TCTAATGGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGAT 780
              ||| ||| ||   || | |||||||||||| |||| |||||||||||||||||||||
Db       1098 TCTTTTGGTATCTCTATCACATATCAAAATGTTCCTGCCGGTTATCGTCCATTTATTGAC 1157

Qy        781 GCTTATATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACT 840
              |||||||||||| |  |||||  |||||| |||   ||  | ||||  |||| ||||||
Db       1158 GCTTATATTTCTCCCTCAGATAATAACCAGTATCAATTGTCGTATAAAAATGACTATACT 1217

Qy        841 TGTGCTGGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAAT 900
              |||| ||    |  |  |||  |  |||||||||||||| |||||||||| ||||||
Db       1218 TGTGTTGATGATTATTGGCAGCATGCACCTTTCACTTTAAAATGGACTGGATATAAGAAT 1277

Qy        901 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1278 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 1337

Qy        961 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1338 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1397

Qy       1021 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1398 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1457

Qy       1081 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1458 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1517

Qy       1141 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGT 1200
             |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db       1518 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACTACAACTCGT 1577

Qy       1201 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1245
             |||||||||||||||||||||||||||||||||||||||||||||
Db       1578 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1622


RESULT 4
AY227439
LOCUS       AY227439                4044 bp    DNA     linear   PLN 30-MAR-2003
DEFINITION  Candida albicans cell-surface adhesin (ALS5) gene, ALS5-small
            allele, complete cds.
ACCESSION   AY227439
VERSION     AY227439.1  GI:29373078
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 4044)
  AUTHORS   Zhao,X., Nuessen,J.A. and Hoyer,L.L.
  TITLE     Analysis of Candida albicans ALS5p function
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4044)
  AUTHORS   Zhao,X., Nuessen,J.A. and Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2003) Veterinary Pathobiology, University of
            Illinois-Urbana, 2522 VMBSB, 2001 S. Lincoln Avenue, Urbana, IL
            61802, USA
FEATURES             Location/Qualifiers
     source          1. .4044
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /strain="SC5314"
                     /db_xref="taxon:5476"
                     /chromosome="6"
     gene            <1. .>4044
                     /gene="ALS5"
                     /allele="small"
     mRNA            <1. .>4044
                     /gene="ALS5"
                     /product="agglutinin-like protein"
     CDS             1. .4044
                     /gene="ALS5"
                     /note="Als5p; cell-surface adhesin"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAO72528.1"
                     /db_xref="GI:29373079"
                     /translation="MLQQFTLLFLYLSFATAKAITGIFNSIDSLTWSNAGNYAFKGPG
                     YPTWNAVLGWSLDGTSANPGDTFILNMPCVFKPTASQKSVDLTADGVKYATCQFYSGE
                     EFTTFSSLKCTVNNNLRSSIKALGTVTLPIAFNVGGTGSSVDLEDSKCFTAGTNTVTF
                     NDGSKKLSIAVNFEKSTVDQSGYLTTSRFMPSLNKIATLYVAPQCENGYTSGTMGFST
                     SYGDVAIDCSNVHIGISKGVNDWNHPVTSESFSYTKSCSSFGISITYQNVPAGYRPFI
                     DAYISPSDNNQYQLSYKNDYTCVDDYWQHAPFTLRWTGYKNSDAGSNGIVIVATTRTV
                     TDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDVPYHTTTTVTSKWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTTTTTQFWSESFTS
                     TTTITNSLKGTDSVIVREPHNPTVTTTEFWSESYATTETITNGPEGTDSVIVREPHNP
                     TVTTTEFWSESYATTETVTNKPEGTDSVIVKEPYNPTVTTTEFWSESYATTETITTGP
                     LGTDSIVIHDPLEESSSTTAIESSDSNISSSAQESSSSVEQSSSIVGLSSSSDIPLSS
                     DMPSSSSTGLTSSESSTVSSYDSDSSSSSELSTFSSSESYSSSISDTTNFWDSSSSDL
                     ESTSITWSSSIDAQSSQSVQSVSNSISTSQETTSSSGEESNTSVTDILVSSDASSILN
                     SDISSYYPSSTISLSDDFPHTIAGEPDSRSSSSIASTVEISSDLVSLTSDPTSSFDSS
                     SSLNSDSSSSPFSDESDISASSSFSTLVAPSFSLSSSSSLSLIYPHYVNSTTYHASES
                     ESSSVASPSMASESANDDTHTLSESTDTTSSIGTDSSTVTFCRRDNGDGCIVTGIPSS
                     SIDSEQTSDVTTTSSFVASSTPTSAEQSITDNPNIDSSQTSASSSTKSSVSVSDTVVN
                     SILLSETSTLSSDDSTSSDTSISSTTNSDTGNINAGSSHTSTASIKESSIQKTGVTLS
                     SSYLSTKLSSTSDITIELITTELITTELTTIEDNEPNTFTSTPSSHSEIFSSDNSVLS
                     KQVDRESTIKTSPTTDVTTVSSLSVHSTEASTATLGENSFSNVASTPSNIATSLRSTS
                     SSSNHATESSGTVKSEASAETIPSPPTSTDNRLSYSTEEAKGITYANSGSTNNLINES
                     QVAAPTDSTSVLIENPVVTSTFDDNSSAAVDQPSKTKSIEESIMNPDSTNETNNGFIA
                     TLSQAQVPSSSIHSELISTTTAKTTDASMNGDSAASNSQPTTLIQQVATSSYNQPLIT
                     TYAGSSSATKHPSWLLKFISVALFPFPL"
BASE COUNT     1242 a    881 c    685 g   1236 t
ORIGIN

  Query Match            82.8%;  Score 1030.6;  DB 8;  Length 4044;
  Best Local Similarity  89.2%;  Pred. No. 1.1e-179;
  Matches 1111;  Conservative    0;  Mismatches  134;  Indels    0;  Gaps    0;

Qy          1 AAGACAATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAAT 60
              ||| | |||||||||| ||||  ||||| || | ||||||||||||||||||||||  |||
Db         52 AAGGCGATCACTGGTATTTTCAATAGTATTGACTCATTAACTTGGTCCAATGCTGGCAAT 111

Qy         61 TATGCTTTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGAT 120
              || |||||||||| |||||||||||||||||||||||||||||| |||||||||| ||||||
Db        112 TACGCTTTCAAAGGACCAGGATACCCAACTTGGAATGCTGTGTTGGGTTGGTCATTAGAT 171

Qy        121 GGTACCAGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATAT 180
              |||||||||||||||||||||||||||||||||||| ||| || ||||||||||||||||||
Db        172 GGTACCAGTGCCAATCCAGGGGATACATTCATATTAAACATGCCATGTGTGTTTAAATTC 231

Qy        181 ACTACTTCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAA 240
              ||| |||| |||| |||||||||||| ||||||||||||||||||||||||||||||||||
Db        232 ACTGCTTCCCAAAAATCTGTTGATTTGACTGCCGATGGTGTTAAATATGCTACTTGTCAA 291

Qy        241 TTTTATTCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCT 300
              |||||||||||||||| || |||||||||||| |||||| |||||| |||||| |   |
Db        292 TTTTATTCTGGTGAAGAGTTTACAACTTTTTCTTCATTAAAATGTACAGTGAACAATAAT 351

Qy        301 TTGAAATCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGT 360
              |||| |||||| |||||||| || ||||| ||||||||||||||||||||||||||||||||
Db        352 TTGAGATCATCTATTAAGGCTTTGGGTACGGTTACTTTACCAATTGCATTCAATGTTGGT 411

Qy        361 GGAACAGGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACA 420
              |||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        412 GGAACAGGGTCATCAGTTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACG 471

Qy        421 GTCACATTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACC 480
              || |||||||||||||   |||| |  |||||||| |||| | ||||||||||||||
Db        472 GTAACATTTAATGATGGCAGTAAAAAGCTCTCAATTGCTGTTAATTTTGAAAAGTCAACA 531

Qy        481 GTTGATCCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACA 540
              ||||||| |||||  ||||||  | ||||||||  |||||| ||||||||||  |  |
Db        532 GTTGATCAAAGTGGGTATTTGACTACTTCCAGATTTATGCCGAGTCTCAATAAAATTGCT 591

Qy        541 ACTCTTTTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCC 600
              ||||||| |||||||||||||||||||| ||||||||||||||||||||||||||| ||||||
Db        592 ACTCTTTATGTGGCACCACAATGTGAAAACGGTTACACATCTGGTACAATGGGATTCTCC 651

Qy        601 AGTAGTAACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGA 660
              |.|||| | || || |||||||||| || |||||| | ||||||||||  ||||||||
Db        652 ACTAGTTATGGGGATGTTGCTATTGACTGTTCAAATGTACATATTGGTATTTCAAAAGGA 711

Qy        661 TTAAATGATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACA 720
              |||||||||||||| |||| ||| | |||||||||||||||||||||||  ||| ||
Db        712 GTAAATGATTGGAATCATCCAGTTACGTCTGAATCATTTAGTTACACTAAAAGCTGTTCA 771

Qy        721 TCTAATGGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGAT 780
              |||  ||| ||    || | |||||||||||| ||||| ||||||||||||||||||||
Db        772 TCTTTTGGTATCTCTATCACATATCAAAATGTTCCTGCCGGTTATCGTCCATTTATTGAC 831

Qy        781 GCTTATATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACT 840
              |||||||||||| |  |||||  ||||||  |||   ||  | ||||  |||||  ||||||
Db        832 GCTTATATTTCTCCCTCAGATAATAACCAGTATCAATTGTCGTATAAAAATGACTATACT 891

Qy        841 TGTGCTGGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAAT 900
```

```
Db        892 TGTGTTGATGATTATTGGCAACATGCACCTTTCACTTTAAGATGGACTGGATATAAGAAT 951

Qy        901 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 960

Db        952 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 1011

Qy        961 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1020

Db       1012 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1071

Qy       1021 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1080

Db       1072 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1131

Qy       1081 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1140

Db       1132 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1191

Qy       1141 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGT 1200

Db       1192 CATACTACCACAACTGTTACAAGTAAATGGACAGGAACAATTACGACGACTACAACTCGT 1251

Qy       1201 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1245

Db       1252 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1296


RESULT 5
AY227440
LOCUS       AY227440               4152 bp    DNA     linear   PLN 30-MAR-2003
DEFINITION  Candida albicans cell-surface adhesin (ALS5) gene, ALS5-large
            allele, complete cds.
ACCESSION   AY227440
VERSION     AY227440.1  GI:29373080
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 4152)
  AUTHORS   Zhao,X., Nuessen,J.A. and Hoyer,L.L.
  TITLE     Analysis of Candida albicans ALS5p function
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4152)
  AUTHORS   Zhao,X., Nuessen,J.A. and Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2003) Veterinary Pathobiology, University of
            Illinois-Urbana, 2522 VMBSB, 2001 S. Lincoln Avenue, Urbana, IL
            61802, USA
FEATURES             Location/Qualifiers
     source          1. .4152
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /strain="SC5314"
                     /db_xref="taxon:5476"
                     /chromosome="6"
     gene            <1. .>4152
                     /gene="ALS5"
                     /allele="large"
     mRNA            <1. .>4152
                     /gene="ALS5"
                     /product="agglutinin-like protein"
     CDS             1. .4152
                     /gene="ALS5"
                     /note="Als5p; cell-surface adhesin"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAO72529.1"
                     /db_xref="GI:29373081"
                     /translation="MIQQFTLLFLYLSFATAKAITGIFNSIDSLTWSNAGNYAFKGPG
                     YPTWNAVLGWSLDGTSANPGDTFILNMPCVFKFTASQKSVDLTADGVKYATCQFYSGE
                     EFTTFSSLKCTVNNNLRSSIKALGTVTLPIAFNVGGTGSSVDLEDSKCFTAGTNTVTF
                     NDGSKKLSIAVNFEKSTVDQSGYLTTSRFMPSLNKIATLYVAPQCENGYTSGTMGFST
                     SYGDVAIDCSNVHIGISKGVNDWNHPVTSESFSYTKSCSSFGISITYQNVPAGYRPFI
                     DAYISPSDNNQYQLSYKNDYTCVDDYWQHAPFTLKWTGYKNSDAGSNGIVIVATTRTV
                     TDSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPLPNPTTTTTQFWSESFTS
                     TTTITNSLKGTDSVIVREPHNPTVTTTEFWSESYATTETITTGPLGTDSVIIKEPHNP
                     TVTTTVFWSESYATTETITTGPLGTDSVIIKEPHNPTVTTTKFWSESYATTETITNKP
                     EGTDSVIVREPHNPTVTTTEFWSESYATTETITTGPLGTDSIVIHDPLEESSSTTAIE
                     SSDSNISSSAQESSSSVEQSSSIVGLSSSSDIPLSSDMPSSSSTGLTSSESSTVSSYD
                     SDSSSSSELSTFSSSESYSSSISDTTNFWDSSSSDLESTSITWSSSIDAQSSQSVQSV
                     SNSISTSQETTSSSGEESNTSVTDILVSSDASSILNSDISSYYPSSTISLSDDFPHTI
                     AGEPDSRSSSSIASTVEISSDLVSLTSDPTSSFDSSSSLNSDSSSSPFSDESDISASS
                     SFSTLVAPSFSLSSSSSLSLIYPHYVNSTTYHASESESSSVASPSMASESANDDTHTL
                     SESTDTTSSIGTDSSTVTFCRRDNGDGCIVTGIPSSSIDSEQTSDVTTTSSFVASSTP
                     TSAEQSITDNPNIDSSQTSASSSTKSSVSVSDTVVNSILLSETSTLSSDDSTSSDTSI
                     SSTTNSDTGNINAGSSHTSTASIKESSIQKTGVTLSSSYLSTKLSSTSDITIELITTE
                     LITTELTTIEDNEPNTFTSTPSSHSEIFSSDNSVLSKQVDRESTIKTSPTTDVTTVSS
                     LSVHSTEASTATLGENSFSNVASTPSNIATSLRSTSSSSNHATESSGTVKSEASAEAI
                     PSPPTSTDNRLSYSTEEAKGITYANSGSTNNLITESQVAAPTDSTSVLIENPVVTSTF
                     DDNSSAAVDQPSKTKSIEESIMNPDSTNETNNGFIATLSQSEAPNSLIHSESISTTMA
                     KTTDASINGDSAASNSQPTTLIQQVATSSYNQPLITTYAGSSSATKHPSWLLKFISVA
                     LFFFL"
BASE COUNT     1277 a    904 c    701 g    1270 t
ORIGIN

  Query Match           82.7%;  Score 1029;  DB 8;  Length 4152;
  Best Local Similarity  89.2%;  Pred. No. 2.1e-179;
  Matches 1110;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy          1 AAGACAATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAAT 60

Db         52 AAAGCGATCACTGGTATTTTCAATAGTATTGACTCATTAACTTGGTCCAATGCTGGCAAT 111

Qy         61 TATGCTTTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGAT 120

Db        112 TACGCTTTCAAAGGACCAGGATACCCAACTTGGAATGCTGTGTTGGGTTGGTCATTAGAT 171

Qy        121 GGTACCAGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATAT 180

Db        172 GGTACCAGTGCCAATCCAGGGGATACATTCATATTAAACATGCCATGTGTGTTTAAATTC 231

Qy        181 ACTACTTCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAA 240

Db        232 ACTGCTTCCCAAAAATCTGTTGATTTGACTGCCGATGGTGTTAAATATGCTACTTGTCAA 291

Qy        241 TTTTATTCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCT 300

Db        292 TTTTATTCTGGTGAAGAGTTTACAACTTTTTCTTCATTAAAATGTACAGTGAACAATAAT 351

Qy        301 TTGAAATCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGT 360

Db        352 TTGAGATCATCTATTAAGGCTTTGGGTACGGTTACTTTACCAATTGCATTCAATGTTGGT 411

Qy        361 GGAACAGGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACA 420

Db        412 GGAACAGGGTCATCAGTTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACG 471

Qy        421 GTCACATTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACC 480

Db        472 GTAACATTTAATGATGGCAGTAAAAAGCTCTCAATTGCTGTTAATTTTGAAAAGTCAACA 531

Qy        481 GTTGATCCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACA 540

Db        532 GTTGATCAAAGTGGGTATTTGACTACTTCCAGATTTATGCCGAGTCTCAATAAAATTGCT 591

Qy        541 ACTCTTTTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCC 600

Db        592 ACTCTTTATGTGGCACCACAATGTGAAAACGGTTACACATCTGGTACAATGGGATTCTCC 651

Qy        601 AGTAGTAACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGA 660

Db        652 ACTAGTTATGGGGATGTTGCTATTGACTGTTCAAATGTACATATTGGTATTTCAAAAGGA 711
```

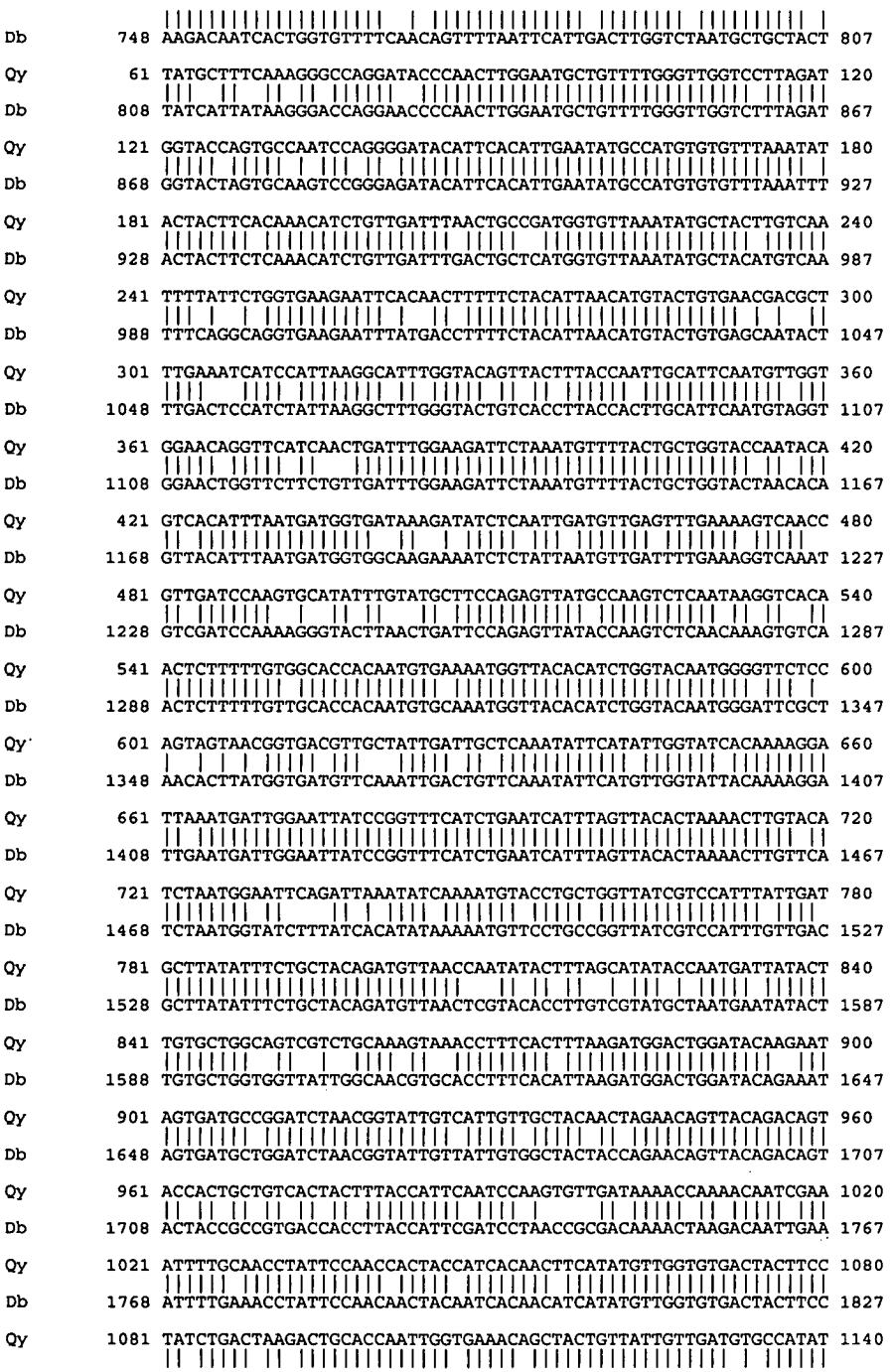
```
Qy        661 TTAAATGATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACA 720
              ||||||||||||||| |||| ||| | ||||||||||||||||||||||||   ||| ||
Db        712 GTAAATGATTGGAATCATCCAGTTACGTCTGAATCATTTAGTTACACTAAAAGCTGTTCA 771

Qy        721 TCTAATGGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGAT 780
              |||  ||| ||    || | ||||||||||||| ||||| ||||||||||||||||||||
Db        772 TCTTTTGGTATCTCTATCACATATCAAAATGTTCCTGCCGGTTATCGTCCATTTATTGAC 831

Qy        781 GCTTATATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACT 840
              |||||||||||| |  |||||  |||||| |||   ||  | ||||  ||||| ||||||
Db        832 GCTTATATTTCTCCCTCAGATAATAACCAGTATCAATTGTCGTATAAAAATGACTATACT 891

Qy        841 TGTGCTGGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAAT 900
              |||| ||    |  |  |||   | |||||||||||||| |||||||||||| ||||||
Db        892 TGTGTTGATGATTATTGGCAGCATGCACCTTTCACTTTAAAATGGACTGGATATAAGAAT 951

Qy        901 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        952 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 1011

Qy        961 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1012 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1071

Qy       1021 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1072 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1131

Qy       1081 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1132 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1191

Qy       1141 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGT 1200
              |||||||||||||||||||||||||||||||||||||| || || || |||||||||
Db       1192 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATTACGACGACTACAACTCGT 1251

Qy       1201 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1245
              |||||||||||||||||||||||||||||||||||||||||||||
Db       1252 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1296
```

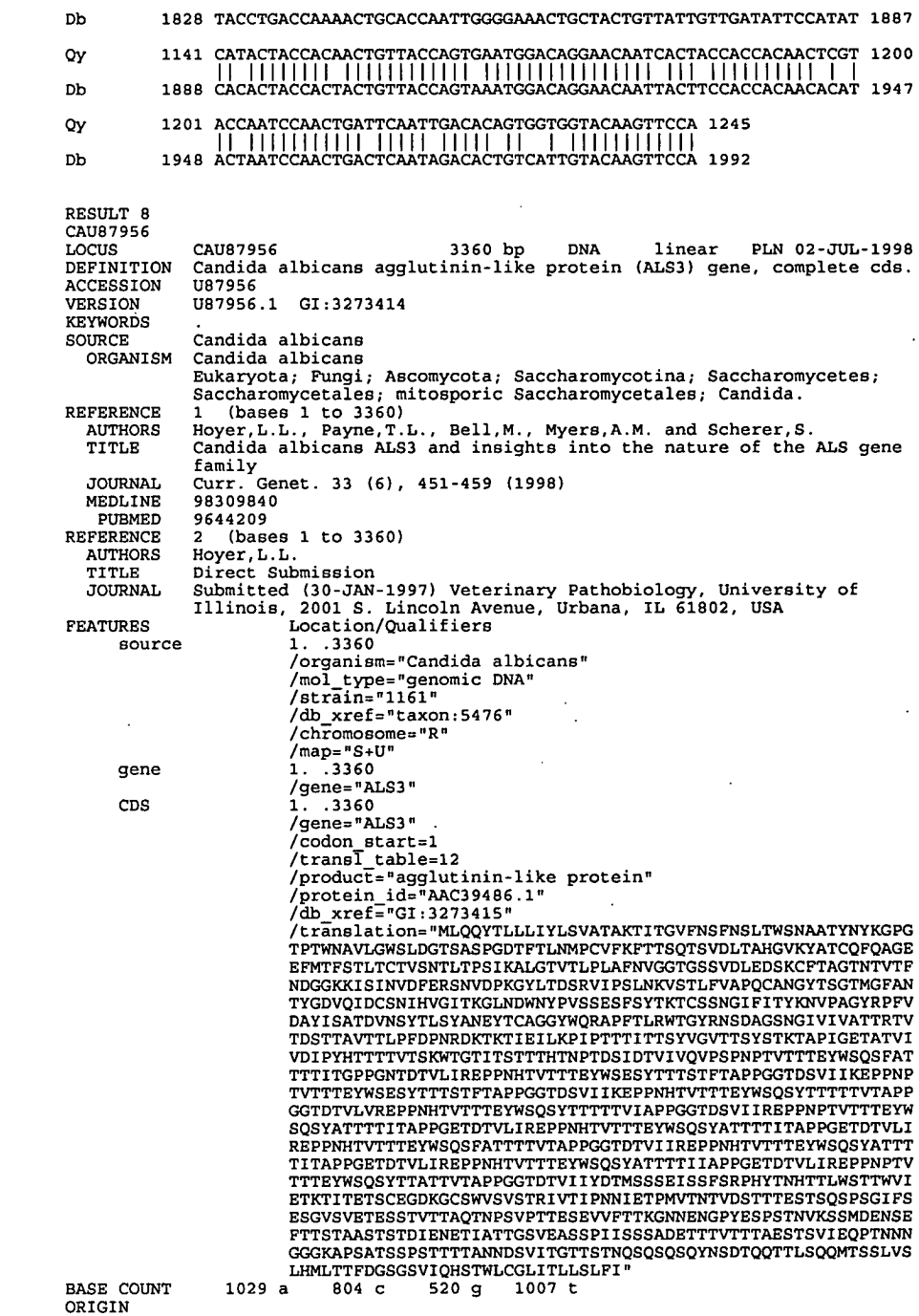
```
RESULT 6
AY223551
LOCUS       AY223551                3144 bp    DNA     linear   PLN 30-MAR-2003
DEFINITION  Candida albicans agglutinin-like sequence 3 (ALS3) gene, ALS3-small
            allele, complete cds.
ACCESSION   AY223551
VERSION     AY223551.1  GI:29373980
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 3144)
  AUTHORS   Zhao,X., Nuessen,J.A., Leng,P., Brown,A.J.P. and Hoyer,L.L.
  TITLE     Candida albicans ALS3 and ALS8 are encoded by a single locus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3144)
  AUTHORS   Zhao,X. and Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2003) Veterinary Pathobiology, University of
            Illinois-Urbana, 2522 VMBSB, 2001 S. Lincoln Avenue, Urbana, IL
            61802, USA
FEATURES             Location/Qualifiers
     source          1. .3144
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /strain="SC5314"
                     /db_xref="taxon:5476"
                     /chromosome="R"
     gene            <1. .>3144
                     /gene="ALS3"
                     /allele="small"
     mRNA            <1. .>3144
                     /gene="ALS3"
                     /product="agglutinin-like sequence 3"
     CDS             1. .3144
                     /gene="ALS3"
                     /note="cell-surface adhesin"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like sequence 3"
                     /protein_id="AAO72958.1"
                     /db_xref="GI:29373981"
                     /translation="MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYNYKGPG
                     TPTWNAVLGWSLDGTSASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGE
                     EFMTFSTLTCTVSNTLTPSIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTF
                     NDGGKKISINVDFERSNVDPKGYLTDSRVIPSLNKVSTLFVAPQCANGYTSGTMGFAN
                     TYGDVQIDCSNIHVGITKGLNDWNYPVSSESFSYTKTCSSNGIFITYKNVPAGYRPFV
                     DAYISATDVNSYTLSYANEYTCAGGYWQRAPFTLRWTGYRNSDAGSNGIVIVATTRTV
                     TDSTTAVTTLPFDPNRDKTKTIEILKPIPTTTIITSYVGVTTSYSTKTAPIGETATVI
                     VDIPYHTTTTVTSKWTGTITSTTTHTNPTDSIDTVIVQVPSPNPTVTTTEYWSQSFAT
                     TTTITGPPGNTDTVLIREPPNHTVTTTEYWSESYTTTSTFTAPPGGTDSVIIKEPPNP
                     TVTTTEYWSESYTTTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSYTTTTTVIAPP
                     GGTDSVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYW
                     SQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSYATTTTITAPPGETDTVLI
                     REPPNHTVTTTEYWSQSYATTTTIIAPPGETDTVLIREPPNPTVTTTEYWSQSYTTAT
                     TVTAPPGGTDTVIIYDTMSSSEISSFSRPHYTNHTTLWSTTWVIETKTITETSCEGDK
                     GCSWVSVSTRIVTIPNNIETPMVTNTVDTTTTESTSQSPSGIFSESGVSVETESSTFT
                     TAQTNPSVPTTSESEVVFTTKGNNGNGPYESPSTNVKSSMDENSEFTTSTAASTSTDIE
                     NETIATTGSVEASSPIISSSADETTTVTTTAESTSVIEQQTNNNGGGNAPSATSTSSP
                     STTTTANSDSVITSTTSTNQSQSQSNSDTQQTTLSQQMTSSLVSLHMLTTFDGSGSVI
                     QHSTWLCGLITLLSLFI"
BASE COUNT      962 a    739 c    491 g    952 t
ORIGIN
```

```
  Query Match             76.2%;  Score 949;  DB 8;  Length 3144;
  Best Local Similarity   85.1%;  Pred. No. 1.1e-164;
  Matches 1060;  Conservative    0;  Mismatches  185;  Indels    0;  Gaps    0;

Qy          1 AAGACAATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAAT 60
              |||||||||||||||||||  | |||||||||||||| ||||||| ||||||||||
Db         52 AAGACAATCACTGGTGTTTTCAACAGTTTTAATTCATTGACTTGGTCTAATGCTGCTACG 111

Qy         61 TATGCTTTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGAT 120
              |||  ||  || || ||||||  ||||||||||||||||||||||||||||| ||||||
Db        112 TATAATTATAAGGGACCAGGAACCCCAACTTGGAATGCTGTTTTGGGTTGGTCTTTAGAT 171

Qy        121 GGTACCAGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATAT 180
              ||||| ||||| | ||| || ||||||||||||||||||||||||||||||||||| |
Db        172 GGTACTAGTGCAAGTCCGGGAGATACATTCACATTGAATATGCCATGTGTGTTTAAATTT 231

Qy        181 ACTACTTCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAA 240
              |||||||| |||||||||||||||| |||||  |||||||||||||||||| ||||||
Db        232 ACTACTTCTCAAACATCTGTTGATTTGACTGCTCATGGTGTTAAATATGCTACATGTCAA 291

Qy        241 TTTTATTCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCT 300
              ||| |  | |||||||||||| |  || |||||||||||||||||||||||||| | |  ||
Db        292 TTTCAGGCAGGTGAAGAATTTATGACCTTTTCTACATTAACATGTACTGTGAGCAATACT 351

Qy        301 TTGAAATCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGT 360
              ||||   |||| |||||||| || ||||| || ||||||||| |||||||||||||| |||
Db        352 TTGACTCCATCTATTAAGGCTTTGGGTACTGTCACTTTACCACTTGCATTCAATGTAGGT 411

Qy        361 GGAACAGGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACA 420
              ||||| |||| ||   |||||||||||||||||||||||||||||||||| || |||
Db        412 GGAACTGGTTCTTCTGTTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACTAACACA 471

Qy        421 GTCACATTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACC 480
              || |||||||||||||||  ||  | |||||||||  |||||| |||||| ||||
Db        472 GTTACATTTAATGATGGTGGCAAGAAAATCTCAATTAATGTTGATTTTGAAAGGTCAAAT 531

Qy        481 GTTGATCCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACA 540
              || |||||||  |  || ||   || |||||||||||| |||||||||| || ||  ||
Db        532 GTCGATCCAAAAGGGTACTTAACTGATTCCAGAGTTATACCAAGTCTCAACAAAGTGTCA 591
```

```
Qy        541 ACTCTTTTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCC 600
              |||||||||||| |||||||||||| |||||||||||||||||||||||||| ||| |
Db        592 ACTCTTTTTGTTGCACCACAATGTGCAAATGGTTACACATCTGGTACAATGGGATTCGCT 651

Qy        601 AGTAGTAACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGA 660
              |  | | | ||||| |||   ||||| || |||||||||||| ||||||| |||||||||
Db        652 AACACTTATGGTGATGTTCAAATTGACTGTTCAAATATTCATGTTGGTATTACAAAAGGA 711

Qy        661 TTAAATGATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACA 720
              || |||||||||||||||||||||||||||||||||||||||||||| ||||||||| ||
Db        712 TTGAATGATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACCAAAACTTGTTCA 771

Qy        721 TCTAATGGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGAT 780
              |||||||| ||    || | |||| ||||||| ||||| |||||||||||||| ||||
Db        772 TCTAATGGTATCTTTATCACATATAAAAATGTTCCTGCCGGTTATCGTCCATTTGTTGAC 831

Qy        781 GCTTATATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACT 840
              |||||||||||||||||||||||||    || || ||  | ||| | ||||| ||||||
Db        832 GCTTATATTTCTGCTACAGATGTTAATTCGTACACCTTGTCGTATGCTAATGAATATACT 891

Qy        841 TGTGCTGGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAAT 900
              ||||||||  ||  |  |||| ||  |||||||| |||||||||||||||||   |||
Db        892 TGTGCTGGTGGTTATTGGCAACGTGCACCTTTCACATTAAGATGGACTGGATACAGAAAT 951

Qy        901 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 960
              ||||||||  |||||||||||||||||  ||||| ||||| || ||||||||||||||||
Db        952 AGTGATGCTGGATCTAACGGTATTGTTATTGTGGCTACTACCAGAACAGTTACAGACAGT 1011

Qy        961 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1020
              || || ||||| || || |||||||||  |||| |     || ||||| || ||||| |||
Db       1012 ACTACCGCTGTGACCACCTTACCATTCGATCCTAACCGCGACAAAACTAAGACAATTGAA 1071

Qy       1021 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1080
              ||||||  |||||||||||||  |||||  |||||||  |||||||||||||||||||||
Db       1072 ATTTTGAAACCTATTCCAACAACTACAATCACAACATCATATGTTGGTGTGACTACTTCC 1131

Qy       1081 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1140
              || ||||| || |||||||||||||| ||||| |||||||||||||||||  | ||||||
Db       1132 TACCTGACCAAAACTGCACCAATTGGGGAAACTGCTACTGTTATTGTTGATATTCCATAT 1191

Qy       1141 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGT 1200
              || |||||||| |||||||||||| |||||||||||||||| ||| |||||||||| | |
Db       1192 CACACTACCACTACTGTTACCAGTAAATGGACAGGAACAATTACTTCCACCACAACACAT 1251

Qy       1201 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1245
              || |||||||||| ||||| ||||| ||  | ||||||||||||
Db       1252 ACTAATCCAACTGACTCAATAGACACTGTCATTGTACAAGTTCCA 1296
```

```
RESULT 7
AF051313
LOCUS       AF051313               4383 bp    DNA     linear   PLN 01-AUG-2000
DEFINITION  Candida albicans agglutinin-like cell surface protein (ALS8) gene,
            complete cds.
ACCESSION   AF051313
VERSION     AF051313.2  GI:9625353
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 4383)
  AUTHORS   Leng,P., Lee,P.R., Wishart,J.A., Wu,H. and Brown,A.J.P.
  TITLE     Sequence of the hypha-specific, agglutinin-like cell surface
            protein, ALS8 from Candida albicans
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4383)
  AUTHORS   Leng,P., Lee,P.R., Wishart,J.A., Wu,H. and Brown,A.J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-1999) Molecular and Cell Biology, University of
            Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen
            AB25 2ZD, UK
COMMENT     On Aug 1, 2000 this sequence version replaced gi:4105850.
FEATURES             Location/Qualifiers
     source          1. .4383
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /strain="ATCC10261"
                     /db_xref="ATCC:10261"
                     /db_xref="taxon:5476"
     gene            566. .>3840
                     /gene="ALS8"
     mRNA            566. .>3840
                     /gene="ALS8"
                     /product="agglutinin-like cell surface protein"
                     /note="minor form"
     protein_bind    574. .579
                     /gene="ALS8"
                     /note="E-box"
                     /bound_moiety="Efg1p"
     TATA_signal     599. .603
                     /gene="ALS8"
     mRNA            612. .>3840
                     /gene="ALS8"
                     /product="agglutinin-like cell surface protein"
                     /note="minor form"
     mRNA            620. .>3840
                     /gene="ALS8"
                     /product="agglutinin-like cell surface protein"
                     /note="major form"
     mRNA            630. .>3840
                     /gene="ALS8"
                     /product="agglutinin-like cell surface protein"
                     /note="minor form"
     CDS             697. .3840
                     /gene="ALS8"
                     /function="involved in cell-cell adhesion"
                     /note="hypha-specific cell surface glycoprotein; similar
                     to Candida albicans strain 1161 agglutinin-like protein
                     encoded by ALS3 gene sequence presented in GenBank
                     Accession Number U87956; contains two fewer internal
                     repeated units than the Candida albicans strain 1161 ALS3
                     gene"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like cell surface protein"
                     /protein_id="AAD02580.1"
                     /db_xref="GI:4105851"
                     /translation="MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYHYKGPG
                     TPTWNAVLGWSLDGTSASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGE
                     EFMTFSTLTCTVSNTLTPSIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTF
                     NDGGKKISINVDFERSNVDPKGYLTDSRVIPSLNKVSTLFVAPQCANGYTSGTMGFAN
                     TYGDVQIDCSNIHVGITKGLNDWNYPVSSESFSYTKTCSSNGIFITYKNVPAGYRPFV
                     DAYISATDVNSYTLSYANEYTCAGGYWQRAPFTLRWTGYRNSDAGSNGIVIVATTRTV
                     TDSTTAVTTLPFDPNRDKTKTIEILKPIPTTTIITTSYVGVTTSYSTKTAPIGETATVI
                     VDIPYHTTTTVTSKWTGTITSTTTHTNPTDSIDTVIVQVPSPNPTVTTTEYWSQSFAT
                     TTTITGPPGNTDTVLIREPPNHTVTTTEYWSESYTTTSTFTAPPGGTDSVIIKEPPNP
                     TVTTTEYWSESYTTTSTFTAPPGGTDSVIIKEPPNPTVTTTEYWSQSYTTTTTVTAPP
                     GGTDTVLVREPPNHTVTTTEYWSQSYTTTTTVIAPPGGTDTVIIREPPNHTVTTTEYW
                     SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTITAPPGETDTVLI
                     REPPNHTVTTTEYWSQSYATTTTVTAPPGGTDTVLIREPPNHTVTTTEYWSQSYATTT
                     TVTAPPGGTDTVIIYDTMSSSEISSFSRPHYTNHTTLWSTTWVIETKTITETSCEGDK
                     GCSWVSVSTRIVTIPNNIETPMVTNTVDSTTTESTSQSPSGIFSESGVSVETESSTVT
                     TAQTNPSVPTTESEVEFTTKGNNGNGPYESPSTHVKSSMDENSEFTTSTAASTSTDIE
                     NATIATTGSVEASSPIISSSADETTTITTTAESTSVIEQPTNNNGGGKAPSATSSPST
                     TTTANNDSVITGTTSTNQSQSQSQSNSDTQQTTLSQQMTSSLVSLHMLTTFDGSGSVI
                     QHSTWLCGLITLLSLFI"
BASE COUNT     1359 a    922 c    705 g   1397 t
ORIGIN

  Query Match           76.2%;  Score 949;  DB 8;  Length 4383;
  Best Local Similarity  85.1%;  Pred. No. 1e-164;
  Matches 1060;  Conservative    0;  Mismatches  185;  Indels    0;  Gaps    0;

Qy          1 AAGACAATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAAT 60
```





RESULT 8
CAU87956
LOCUS CAU87956 3360 bp DNA linear PLN 02-JUL-1998
DEFINITION Candida albicans agglutinin-like protein (ALS3) gene, complete cds.
ACCESSION U87956
VERSION U87956.1 GI:3273414
KEYWORDS .
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 3360)
AUTHORS Hoyer,L.L., Payne,T.L., Bell,M., Myers,A.M. and Scherer,S.
TITLE Candida albicans ALS3 and insights into the nature of the ALS gene family
JOURNAL Curr. Genet. 33 (6), 451-459 (1998)
MEDLINE 98309840
PUBMED 9644209
REFERENCE 2 (bases 1 to 3360)
AUTHORS Hoyer,L.L.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-1997) Veterinary Pathobiology, University of Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES Location/Qualifiers
source 1. .3360
/organism="Candida albicans"
/mol_type="genomic DNA"
/strain="1161"
/db_xref="taxon:5476"
/chromosome="R"
/map="S+U"
gene 1. .3360
/gene="ALS3"
CDS 1. .3360
/gene="ALS3"
/codon_start=1
/transl_table=12
/product="agglutinin-like protein"
/protein_id="AAC39486.1"
/db_xref="GI:3273415"
/translation="MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYNYKGPG
TPTWNAVLGWSLDGTSASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGE
EFMTFSTLTCTVSNTLTPSIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTF
NDGGKKISINVDFERSNVDPKGYLTDSRVIPSLNKVSTLFVAPQCANGYTSGTMGFAN
TYGDVQIDCSNIHVGITKGLNDWNYPVSSESFSYTKTCSSNGIFITYKNVPAGYRPFV
DAYISATDVNSYTLSYANEYTCAGGYWQRAPFTLRWTGYRNSDAGSNGIVIVATTRTV
TDSTTAVTTLPFDPNRDKTKTIEILKPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
VDIPYHTTTTVTSKWTGTITSTTTHTNPTDSIDTVIVQVPSPNPTVTTTEYWSQSFAT
TTTITGPPGNTDTVLIREPPNHTVTTTEYWSESYTTTSTFTAPPGGTDSVIIKEPPNP
TVTTTEYWSESYTTTSTFTAPPGGTDSVIIKEPPNHTVTTTEYWSQSYTTTTTVTAPP
GGTDTVLVREPPNHTVTTTEYWSQSYTTTTTVIAPPGGTDSVIIREPPNPTVTTTEYW
SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTITAPPGETDTVLI
REPPNHTVTTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSYATTT
TITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTTTIIAPPGETDTVLIREPPNPTV
TTTEYWSQSYTTATTVTAPPGGTDTVIIYDTMSSSEISSFSRPHYTNHTTLWSTTWVI
ETKTITETSCEGDKGCSWVSVSTRIVTIPNNIETPMVTNTVDSTTTESTSQSPSGIFS
ESGVSVETESSTVTTAQTNPSVPTTESEVVFTTKGNNENGPYESPSTNVKSSMDENSE
FTTSTAASTSTDIENETIATTGSVEASSPIISSSADETTTVTTTAESTSVIEQPTNNN
GGGKAPSATSSPSTTTTANNDSVITGTTSTNQSQSQSQYNSDTQQTTLSQQMTSSLVS
LHMLTTFDGSGSVIQHSTWLCGLITLLSLFI"
BASE COUNT 1029 a 804 c 520 g 1007 t
ORIGIN

Query Match 76.1%; Score 947.4; DB 8; Length 3360;
Best Local Similarity 85.1%; Pred. No. 2.2e-164;
Matches 1059; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

```
Qy          1 AAGACAATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAAT 60
              |||||||||||||||||||||  | ||||||||||||||| ||||||| |||||||||| |
Db         52 AAGACAATCACTGGTGTTTTCAACAGTTTTAATTCATTGACTTGGTCTAATGCTGCTACT 111

Qy         61 TATGCTTTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGAT 120
              |||  ||  || || ||||||  |||||||||||||||||||||||||||||||| ||||||
Db        112 TATAATTATAAGGGACCAGGAACCCCAACTTGGAATGCTGTTTTGGGTTGGTCTTTAGAT 171

Qy        121 GGTACCAGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATAT 180
              |||||  |||||  | ||| || |||||||||||||||||||||||||||||||||||| |
Db        172 GGTACTAGTGCAAGTCCGGGAGATACATTCACATTGAATATGCCATGTGTGTTTAAATTT 231

Qy        181 ACTACTTCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAA 240
              ||||||||  |||||||||||||||||||  |||||  |||||||||||||||||||  ||||||
Db        232 ACTACTTCTCAAACATCTGTTGATTTGACTGCTCATGGTGTTAAATATGCTACATGTCAA 291

Qy        241 TTTTATTCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCT 300
              |||  |  | ||||||||||| |  || |||||||||||||||||||||||| | |  ||
Db        292 TTTCAGGCAGGTGAAGAATTTATGACCTTTTCTACATTAACATGTACTGTGAGCAATACT 351

Qy        301 TTGAAATCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGT 360
              ||||   |||| |||||||| || ||||| || || |||||| |||||||||||||  |||
Db        352 TTGACTCCATCTATTAAGGCTTTGGGTACTGTCACCTTACCACTTGCATTCAATGTAGGT 411

Qy        361 GGAACAGGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACA 420
              |||||  |||||  ||   |||||||||||||||||||||||||||||||||||| || |||
Db        412 GGAACTGGTTCTTCTGTTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACTAACACA 471

Qy        421 GTCACATTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACC 480
              || |||||||||||||||  ||  | ||||| ||| |||||||  |||||||  |||||
Db        472 GTTACATTTAATGATGGTGGCAAGAAAATCTCTATTAATGTTGATTTTGAAAGGTCAAAT 531

Qy        481 GTTGATCCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACA 540
              || |||||||  |  || ||   || ||||||||||| |||||||||| || ||  ||
Db        532 GTCGATCCAAAAGGGTACTTAACTGATTCCAGAGTTATACCAAGTCTCAACAAAGTGTCA 591

Qy        541 ACTCTTTTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCC 600
              ||||||||||| |||||||||||| |||||||||||||||||||||||||||||| ||| |
Db        592 ACTCTTTTTGTTGCACCACAATGTGCAAATGGTTACACATCTGGTACAATGGGATTCGCT 651

Qy        601 AGTAGTAACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGA 660
              |  | | | |||||  |||    ||||| || |||||||||||  |||||||  |||||||||
Db        652 AACACTTATGGTGATGTTCAAATTGACTGTTCAAATATTCATGTTGGTATTACAAAAGGA 711

Qy        661 TTAAATGATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACA 720
              || |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||| ||
Db        712 TTGAATGATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACCAAAACTTGTTCA 771

Qy        721 TCTAATGGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGAT 780
              |||||||| ||    || | |||| ||||||| ||||| ||||||||||||||| ||||
Db        772 TCTAATGGTATCTTTATCACATATAAAAATGTTCCTGCCGGTTATCGTCCATTTGTTGAC 831

Qy        781 GCTTATATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACT 840
              |||||||||||||||||||||||||    || || ||  | ||| | |||||  ||||||
Db        832 GCTTATATTTCTGCTACAGATGTTAATTCGTACACCTTGTCGTATGCTAATGAATATACT 891

Qy        841 TGTGCTGGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAAT 900
              ||||||||  ||  |  |||| ||  |||||||| ||||||||||||||||||  |||
Db        892 TGTGCTGGTGGTTATTGGCAACGTGCACCTTTCACATTAAGATGGACTGGATACAGAAAT 951

Qy        901 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 960
              ||||||||  |||||||||||||||||||  |||||  |||||  || ||||||||||||||||||
Db        952 AGTGATGCTGGATCTAACGGTATTGTTATTGTGGCTACTACCAGAACAGTTACAGACAGT 1011

Qy        961 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1020
              || || ||||| || || |||||||||  |||| |     || ||||| || |||||  |||
Db       1012 ACTACCGCTGTGACCACCTTACCATTCGATCCTAACCGCGACAAAACTAAGACAATTGAA 1071

Qy       1021 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1080
              |||||| |||||||||||||| ||||| |||||||| |||||||||||||||||||||||||
Db       1072 ATTTTGAAACCTATTCCAACAACTACAATCACAACATCATATGTTGGTGTGACTACTTCC 1131

Qy       1081 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1140
              || ||||| || |||||||||||||| ||||| |||||||||||||||||| | ||||||
Db       1132 TACCTGACCAAAACTGCACCAATTGGGGAAACTGCTACTGTTATTGTTGATATTCCATAT 1191

Qy       1141 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGT 1200
              || |||||||| |||||||||||| |||||||||||||||| ||| |||||||||| | |
Db       1192 CACACTACCACTACTGTTACCAGTAAATGGACAGGAACAATTACTTCCACCACAACACAT 1251

Qy       1201 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1245
              || |||||||||||| ||||| ||||| ||  | ||||||||||||
Db       1252 ACTAATCCAACTGACTCAATAGACACTGTCATTGTACAAGTTCCA 1296
```

RESULT 9
AY223552

```
LOCUS       AY223552               3468 bp    DNA     linear   PLN 30-MAR-2003
DEFINITION  Candida albicans agglutinin-like sequence 3 (ALS3) gene, ALS3-large
            allele, complete cds.
ACCESSION   AY223552
VERSION     AY223552.1  GI:29373982
KEYWORDS    .
SOURCE      Candida albicans.
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 3468)
  AUTHORS   Zhao,X., Nuessen,J.A., Leng,P., Brown,A.J.P. and Hoyer,L.L.
  TITLE     Candida albicans ALS3 and ALS8 are encoded by a single locus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3468)
  AUTHORS   Zhao,X. and Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2003) Veterinary Pathobiology, University of
            Illinois-Urbana, 2522 VMBSB, 2001 S. Lincoln Avenue, Urbana, IL
            61802, USA
FEATURES             Location/Qualifiers
     source          1. .3468
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /strain="SC5314"
                     /db_xref="taxon:5476"
                     /chromosome="R"
     gene            <1. .>3468
                     /gene="ALS3"
                     /allele="large"
     mRNA            <1. .>3468
                     /gene="ALS3"
                     /product="agglutinin-like sequence 3"
     CDS             1. .3468
                     /gene="ALS3"
                     /note="cell-surface adhesin"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like sequence 3"
                     /protein_id="AAO72959.1"
                     /db_xref="GI:29373983"
                     /translation="MLQQYTLLLIYLSVATAKTITGVFNSFNSLTWSNAATYNYKGPG
                     TPTWNAVLGWSLDGTSASPGDTFTLNMPCVFKFTTSQTSVDLTAHGVKYATCQFQAGE
                     EFMTFSTLTCTVSNTLTPSIKALGTVTLPLAFNVGGTGSSVDLEDSKCFTAGTNTVTF
                     NDGGKKISINVDFERSNVDPKGYLTDSRVIPSLNKVSTLFVAPQCANGYTSGTMGFAN
                     TYGDVQIDCSNIHVGITKGLNDWNYPVSSESFSYTKTCSSNGIFITYKNVPAGYRPFV
                     DAYISATDVNSYTLSYANEYTCAGGYWQRAPFTLRWTGYRNSDAGSNGIVIVATTRTV
                     TDSTTAVTTLPFPDPNRDKTKTIEILKPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDIPYHTTTTVTSKWTGTITSTTTHTNPTDSIDTVIVQVPSPNPTVTTTEYWSQSFAT
                     TTTITGPPGNTDTVLIREPPNHTVTTTEYWSESYTTTSTFTAPPGGTDSVIIKEPPNP
                     TVTTTEYWSESYTTTTTVTAPPGGTDTVIIREPPNHTVTTTEYWSQSYTTTTTVIAPP
                     GGTDSVIIREPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYW
                     SQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYTTTTTVIAPPGETDSVII
                     KEPPNPTVTTTEYWSQSYATTTTITAPPGETDTVLIREPPNHTVTTTEYWSQSYATTT
```

```
                     TITAPPGETDTVLIREPPNHTVTTTEYWSQSFATTTTVTAPPGGTDTVIIREPPNHTV
                     TTTEYWSQSFATTTTVTAPPGGTDTVLIREPPNPTVTTTEYWSQPYTTTTTVIAPPGG
                     TDTVIIYDTMSSSEISSFSRPHYTNHTTLWSTTWVIETKTITETSCEGDKGCSWVSVS
                     TRIVTIPNNIETPMVTNTVDSTTTESTSQSPSGIFSESGVSVETESSTVTTAQTNPSV
                     PTTESEVVFTTKGNNGNGPYESPSTNVKSSMDENSEFTTSTAASTSTDIENETIATTG
                     SVEASSPIISSSADETTTVTTTAESTSVIEQPTNNNGGGNAPSATSTSSPSTTTTANS
                     DSVITSTTSTNQSQSQSNSDTQQTTLSQQMTSSLVSLHMLTTFDGSGSVIQHSTWLCG
                     LITLLSLFI"
BASE COUNT     1069 a    841 c    533 g   1025 t
ORIGIN

  Query Match           76.0%;  Score 945.8;  DB 8;  Length 3468;
  Best Local Similarity   85.0%;  Pred. No. 4.2e-164;
  Matches 1058;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy          1 AAGACAATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAAT 60
              |||||||||||||||||||  | ||||||||||||| |||||||| ||||||||||
Db         52 AAGACAATCACTGGTGTTTTCAACAGTTTTAATTCATTGACTTGGTCTAATGCTGCTACG 111

Qy         61 TATGCTTTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGAT 120
              |||  ||  || || ||||||  ||||||||||||||||||||||||||||||| ||||||
Db        112 TATAATTATAAGGGACCAGGAACCCCAACTTGGAATGCTGTTTTGGGTTGGTCTTTAGAT 171

Qy        121 GGTACCAGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATAT 180
              ||||| ||||| | ||| || ||||||||||||||||||||||||||||||||||||| |
Db        172 GGTACTAGTGCAAGTCCGGGAGATACATTCACATTGAATATGCCATGTGTGTTTAAATTT 231

Qy        181 ACTACTTCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAA 240
              |||||||| |||||||||||||||| |||||  |||||||||||||||||||||||||||||
Db        232 ACTACTTCTCAAACATCTGTTGATTTGACTGCTCATGGTGTTAAATATGCTACATGTCAA 291

Qy        241 TTTTATTCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCT 300
              ||| |  | |||||||||||  |  || |||||||||||||||||||||||| | |  ||
Db        292 TTTCAGGCAGGTGAAGAATTTATGACCTTTTCTACATTAACATGTACTGTGAGCAATACT 351

Qy        301 TTGAAATCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGT 360
              ||||  |||| |||||||| || ||||| || |||||||| |||||||||||||| |||
Db        352 TTGACTCCATCTATTAAGGCTTTGGGTACTGTCACTTTACCACTTGCATTCAATGTAGGT 411

Qy        361 GGAACAGGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACA 420
              ||||| ||||| ||   |||||||||||||||||||||||||||||||||||||| ||| |||
Db        412 GGAACTGGTTCTTCTGTTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACTAACACA 471

Qy        421 GTCACATTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACC 480
              || |||||||||||||||  ||  | ||||||||| ||||||| ||||||| |||||
Db        472 GTTACATTTAATGATGGTGGCAAGAAAATCTCAATTAATGTTGATTTTGAAAGGTCAAAT 531

Qy        481 GTTGATCCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACA 540
              || |||||||  |  || ||   || |||||||||||| |||||||||| || ||  ||
Db        532 GTCGATCCAAAAGGGTACTTAACTGATTCCAGAGTTATACCAAGTCTCAACAAAGTGTCA 591

Qy        541 ACTCTTTTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCC 600
              |||||||||||  |||||||||||||||||||||||||||||||||||||||||||| ||| |
Db        592 ACTCTTTTTGTTGCACCACAATGTGCAAATGGTTACACATCTGGTACAATGGGATTCGCT 651

Qy        601 AGTAGTAACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGA 660
              |  | | | ||||| |||   ||||| || |||||||||||| ||||||| |||||||||
Db        652 AACACTTATGGTGATGTTCAAATTGACTGTTCAAATATTCATGTTGGTATTACAAAAGGA 711

Qy        661 TTAAATGATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACA 720
              || ||||||||||||||||||||||||||||||||||||||||||||| ||||||||| ||
Db        712 TTGAATGATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACCAAAACTTGTTCA 771

Qy        721 TCTAATGGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGAT 780
              |||||||| ||    || | |||| |||| || ||||| ||||||||||||| ||||
Db        772 TCTAATGGTATCTTTATCACATATAAAAACGTTCCTGCCGGTTATCGTCCATTTGTTGAC 831

Qy        781 GCTTATATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACT 840
              |||||||||||||||||||||||||    || || ||  | ||| | ||||| ||||||
Db        832 GCTTATATTTCTGCTACAGATGTTAATTCGTACACCTTGTCGTATGCTAATGAATATACT 891

Qy        841 TGTGCTGGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAAT 900
              ||||||||  ||  |  |||| ||  ||||||||| ||||||||||||||||||  |||

Db        892 TGTGCTGGTGGTTATTGGCAACGTGCACCTTTCACATTAAGATGGACTGGATACAGAAAT 951

Qy        901 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 960
              |||||||| |||||||||||||||| ||||| ||||| || ||||||||||||||||||
Db        952 AGTGATGCTGGATCTAACGGTATTGTTATTGTGGCTACTACCAGAACAGTTACAGACAGT 1011

Qy        961 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1020
              || || || || || || ||||||||| |||| |      || ||||| || ||||| |||
Db       1012 ACTACCGCCGTGACCACCTTACCATTCGATCCTAACCGCGACAAAACTAAGACAATTGAA 1071

Qy       1021 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1080
              |||||| |||||||||||||| ||||| |||||||| ||||||||||||||||||||||||
Db       1072 ATTTTGAAACCTATTCCAACAACTACAATCACAACATCATATGTTGGTGTGACTACTTCC 1131

Qy       1081 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1140
              || ||||| || |||||||||||||| ||||| |||||||||||||||||| | ||||||
Db       1132 TACCTGACCAAAACTGCACCAATTGGGGAAACTGCTACTGTTATTGTTGATATTCCATAT 1191

Qy       1141 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGT 1200
              || |||||||| |||||||||||| |||||||||||||||| ||| |||||||||| | |
Db       1192 CACACTACCACTACTGTTACCAGTAAATGGACAGGAACAATTACTTCCACCACAACACAT 1251

Qy       1201 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1245
              || |||||||||| ||||| ||||| ||  | ||||||||||||
Db       1252 ACTAATCCAACTGACTCAATAGACACTGTCATTGTACAAGTTCCA 1296


RESULT 10
CA1ALS1
LOCUS       CA1ALS1                 1404 bp    DNA     linear   PLN 16-OCT-1998
DEFINITION  Candida albicans agglutinin-like protein (ALS2) gene, 5' partial
            cds.
ACCESSION   AF024580
VERSION     AF024580.1  GI:3598667
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 1404)
  AUTHORS   Hoyer,L.L., Payne,T.L. and Hecht,J.E.
  TITLE     Identification of Candida albicans ALS2 and ALS4 and localization
            of als proteins to the fungal cell surface
  JOURNAL   J. Bacteriol. 180 (20), 5334-5343 (1998)
  MEDLINE   98440424
   PUBMED   9765564
REFERENCE   2  (bases 1 to 1404)
  AUTHORS   Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1997) Veterinary Pathobiology, University of
            Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES             Location/Qualifiers
     source          1. .1404
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="6"
                     /map="6; C"
     CDS             1. .>1404
                     /gene="ALS2"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAC64235.1"
                     /db_xref="GI:3598670"
                     /translation="MLLQFLLLSLCVSVATAKVITGVFNSFDSLTWTRAGNYAYKGPN
                     RPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKFITDQTSVDLTAEGVKYATCQFYSGE
                     EFTTFSSLKCTVSNTLTSSIKALGTVTLPISFNVGGTGSSVDLESSQCFKAGTNTVTF
                     NDGDKKISIDVDFEKTNEDASGYFIASRLIPSINKVSITYVAPQCANGYTSGAMGFIV
                     LTGDTTIDCSNVHVGITKGLNDWNFPVSSDSLSYNKTCSSTGISITYENVPAGYRPFF
                     DVYTSVSGQNRQLRYTNDYACVGSSLQSKPFNLRLRGYNNSEANSNGFVIVATTRTVT
```

```
                     DSTTAVTTLPFNPSVDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVIV
                     DVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPSPNPTVTTTEYWSQSYATT
                     TTVTAPPGGTDSVIIREP"
     repeat_region   1297. .1404
                     /note="first of multiple copies of imperfect 108 bp
                     repeat"
                     /rpt_type=tandem
BASE COUNT      422 a    268 c    253 g    461 t
ORIGIN
```

Query Match 71.9%; Score 895; DB 8; Length 1404;
Best Local Similarity 83.2%; Pred. No. 1.1e-154;
Matches 1031; Conservative 0; Mismatches 205; Indels 3; Gaps 1;

```
Qy          7 ATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCT 66
Db         58 ATTACGGGTGTTTTCAATAGTTTTGATTCGTTGACATGGACAAGAGCTGGTAATTATGCT 117

Qy         67 TTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACC 126
Db        118 TATAAGGGCCCAAATAGACCAACTTGGAATGCTGTTTTGGGCTGGTCTTTAGATGGTACT 177

Qy        127 AGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACT 186
Db        178 AGTGCAAATCCAGGAGACACATTCACATTGAATATGCCATGTGTTTTTAAATTTATTACC 237

Qy        187 TCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTAT 246
Db        238 GATCAAACATCTGTTGATTTGACTGCTGAAGGTGTTAAATATGCTACATGTCAGTTTTAT 297

Qy        247 TCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAA 306
Db        298 TCAGGTGAAGAATTTACAACATTTTCTTCATTAAAATGTACTGTGAGCAATACTTTAACA 357

Qy        307 TCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACA 366
Db        358 TCATCTATTAAGGCTTTGGGTACGGTTACTTTACCAATTTCATTTAATGTTGGTGGAACA 417

Qy        367 GGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACA 426
Db        418 GGTTCATCGGTTGATTTGGAAAGTTCTCAATGTTTTAAGGCTGGCACCAACACAGTTACT 477

Qy        427 TTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGAT 486
Db        478 TTTAATGATGGTGATAAAAAAATCTCAATTGACGTTGATTTTGAGAAAACAAACGAAGAT 537

Qy        487 CCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTT 546
Db        538 GCAAGTGGATATTTCATAGCGTCAAGACTTATTCCAAGTATTAACAAAGTTTCAATCACT 597

Qy        547 TTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGT 606
Db        598 TATGTGGCACCACAATGTGCAAATGGCTACACATCTGGTGCAATGGGGTTCATAGTTCTC 657

Qy        607 AACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAAT 666
Db        658 ACTGGTGACACTACTATTGACTGTTCAAATGTTCATGTTGGTATTACAAAGGGATTAAAT 717

Qy        667 GATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAAT 726
Db        718 GATTGGAATTTTCCGGTATCGTCTGATTCATTAAGTTACAATAAAACTTGTTCATCTACA 777

Qy        727 GGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTAT 786
Db        778 GGTATTTCTATCACATATGAAAATGTCCCCGCTGGTTATCGTCCATTTTTTGACGTATAT 837

Qy        787 ATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCT 846
Db        838 ACTCTGGTGTCAGGCCAGAACAGACA---ATTAAGATATACTAATGATTATGCCTGTGTT 894

Qy        847 GGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGAT 906
Db        895 GGTAGTTCCTTACAAAGTAAGCCGTTCAATTTAAGATTGAGAGGATACAATAATAGTGAA 954
```

```
Qy        907 GCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACT 966
Db        955 GCTAATTCTAACGGTTTTGTCATTGTTGCTACAACCCGAACAGTTACTGACAGTACTACT 1014

Qy        967 GCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTG 1026
Db       1015 GCTGTCACTACTTTACCTTTTAATCCAAGTGTTGACAAAACCAAAACAATCGAAATTTTG 1074

Qy       1027 CAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTG 1086
Db       1075 CAACCTATTCCAACAACCACCATCACAACTTCATATGTTGGTGTGACTACTTCCTACCTG 1134

Qy       1087 ACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACT 1146
Db       1135 ACTAAAACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACT 1194

Qy       1147 ACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAAT 1206
Db       1195 ACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACTACAACTCGTACCAAT 1254

Qy       1207 CCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1245
Db       1255 CCAACTGATTCTATAGATACTGTCGTTGTGCAAGTTCCA 1293
```

```
RESULT 11
CANALS2S1
LOCUS       CANALS2S1               1404 bp    DNA     linear   PLN 16-OCT-1998
DEFINITION  Candida albicans agglutinin-like protein (ALS2) gene, 5' partial
            cds.
ACCESSION   AF024582
VERSION     AF024582.1  GI:3598672
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 1404)
  AUTHORS   Hoyer,L.L., Payne,T.L. and Hecht,J.E.
  TITLE     Identification of Candida albicans ALS2 and ALS4 and localization
            of als proteins to the fungal cell surface
  JOURNAL   J. Bacteriol. 180 (20), 5334-5343 (1998)
  MEDLINE   98440424
   PUBMED   9765564
REFERENCE   2  (bases 1 to 1404)
  AUTHORS   Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1997) Veterinary Pathobiology, University of
            Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES             Location/Qualifiers
     source          1. .1404
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="6"
                     /map="6C"
     gene            1. .>1404
                     /gene="ALS2"
                     /note="agglutinin-like sequence"
                     /allele="ALS2-2"
     CDS             1. .>1404
                     /gene="ALS2"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAC64237.1"
                     /db_xref="GI:3598675"
                     /translation="MLLQFLLLSLCVSVATAKVITGVFNSFDSLTWTRAGNYAYKGPN
                     RPTWNAVLGWSLDGTSANPGDTFTLNMPCVFKFITDQTSVDLTAEGVKYATCQFYSGE
                     EFTTFSSLKCTVSNTLTSSIKALGTVTLPISFNVGGTGSSVDLESSQCFKAGTNTVTF
```

```
                     NDGDKKISIDVDFEKTNEDASGFFIASRLIPSINKVSITYVAPQCANGYTSGAMGFIV
                     LTGDTTIDCSNVHVGITKGLNDWNFPVSSDSLSYNKTCSSTGISITYENVPAGYRPFF
                     DVYTSVSGQNRQLRYTNDYACVGSSLQSKPFNLRLRGYNNSEANSNGFVIVATTRTVT
                     DSTTAVTTLPFNPSVDKTKTIEILQPIPTTTIITSYVGVTTSYSTKTAPIGETATVIV
                     DVPYHTTTTVTSEWTGTITATTTRTNPTDSIDTVVVQVPSPNPTVSTTEYWSQSYATT
                     TTVTAPPGGTDSVIIREP"
     repeat_region   1297. .1404
                     /note="first of multiple copies of 108bp imperfect repeat"
                     /rpt_type=tandem
BASE COUNT      421 a    266 c    255 g    462 t
ORIGIN

  Query Match             71.5%;  Score 890.2;  DB 8;  Length 1404;
  Best Local Similarity   83.0%;  Pred. No. 8.7e-154;
  Matches 1028;  Conservative    0;  Mismatches  208;  Indels    3;  Gaps    1;

Qy          7 ATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCT 66
              || || ||||||||  ||||||||| |||| || || ||| | |  |||| ||||||||||
Db         58 ATTACGGGTGTTTTCAATAGTTTTGATTCGTTGACATGGACAAGAGCTGGTAATTATGCT 117

Qy         67 TTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACC 126
              |  || || |||      ||||||||||||||||||||||| ||||| ||||||||||
Db        118 TATAAGGGCCCAAATAGACCAACTTGGAATGCTGTTTTGGGCTGGTCTTTAGATGGTACT 177

Qy        127 AGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACT 186
              ||||| |||||||| || ||||||||||||||||||||||||| ||||||| || |||
Db        178 AGTGCAAATCCAGGAGACACATTCACATTGAATATGCCATGTGTTTTTAAATTTATTACC 237

Qy        187 TCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTAT 246
              ||||||||||||||||| ||||| || |||||||||||||||| ||||| ||||||
Db        238 GATCAAACATCTGTTGATTTGACTGCTGAAGGTGTTAAATATGCTACATGTCAGTTTTAT 297

Qy        247 TCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAA 306
              || |||||||||| ||||| |||||| |||||| |||||||||| | |  |||| | |
Db        298 TCAGGTGAAGAATTTACAACATTTTCTTCATTAAAATGTACTGTGAGCAATACTTTAACA 357

Qy        307 TCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACA 366
              ||||| |||||||| || ||||| |||||||||||||||| |||| ||||||||||||||
Db        358 TCATCTATTAAGGCTTTGGGTACGGTTACTTTACCAATTTCATTTAATGTTGGTGGAACA 417

Qy        367 GGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACA 426
              ||||||||   ||||||||||  |||| |||||||||  ||||| ||||| ||||| ||
Db        418 GGTTCATCGGTTGATTTGGAAAGTTCTCAATGTTTTAAGGCTGGCACCAACACAGTTACT 477

Qy        427 TTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGAT 486
              |||||||||||||||| | || |||||||| ||||| ||||| ||  ||| ||  |||
Db        478 TTTAATGATGGTGATAAAAAAATTTCAATTGACGTTGATTTTGAGAAAACAAACGAAGAT 537

Qy        487 CCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTT 546
              |||||| || |||     || || ||| |||| |||||| | || || ||  |||     |
Db        538 GCAAGTGGATTTTTCATAGCGTCAAGACTTATTCCAAGTATTAACAAAGTTTCAATCACT 597

Qy        547 TTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGT 606
              | |||||||||||||||| |||||| |||||||||||| ||||||||||      |
Db        598 TATGTGGCACCACAATGTGCAAATGGCTACACATCTGGTGCAATGGGGTTCATAGTTCTC 657

Qy        607 AACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAAT 666
              |  ||||||  | ||||||| || |||||| ||||| |||||||||||| |||||||||
Db        658 ACTGGTGACACTACTATTGACTGTTCAAATGTTCATGTTGGTATCACAAAGGGATTAAAT 717

Qy        667 GATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAAT 726
              |||||||||| |||||| || ||||| ||||| ||||||| |||||||||| ||||||
Db        718 GATTGGAATTTTCCGGTATCGTCTGATTCATTAAGTTACAATAAAACTTGTTCATCTACA 777

Qy        727 GGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTAT 786
              || |||   || | |||| ||||||| || |||||||||||||||||| |||| |  |||
Db        778 GGTATTTCTATCACATATGAAAATGTCCCCGCTGGTTATCGTCCATTTTTTGACGTATAT 837

Qy        787 ATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCT 846
              | |   |  |||     |||  | |    |||  |||||| ||||||||| | |||| |
Db        838 ACTCTGGTGTCAGGCCAGAACAGACA---ATTAAGATATACTAATGATTATGCCTGTGTT 894

Qy        847 GGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGAT 906
              || |||    | |||||||| || |||| |||||||| ||  |||||||| ||||||||
Db        895 GGTAGTTCCTTACAAAGTAAGCCGTTCAATTTAAGATTGAGAGGATACAATAATAGTGAA 954

Qy        907 GCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACT 966
              ||    ||||||||| ||||||||||||||||||||  |||||||||  |||||||| |||
Db        955 GCTAATTCTAACGGTTTTGTCATTGTTGCTACAACCCGAACAGTTACTGACAGTACTACT 1014

Qy        967 GCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTG 1026
              |||||||||||||||| || ||||||||||||| ||||||||||||||||||||||||
Db       1015 GCTGTCACTACTTTACCTTTTAATCCAAGTGTTGACAAAACCAAAACAATCGAAATTTTG 1074

Qy       1027 CAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTG 1086
              ||||||||||||| || || |||||||||||||||||||||||||||||||||| |||
Db       1075 CAACCTATTCCAACAACCACTATCACAACTTCATATGTTGGTGTGACTACTTCCTACCTG 1134

Qy       1087 ACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACT 1146
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1135 ACTAAAACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACT 1194

Qy       1147 ACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAAT 1206
              |||||||||||||||||||||||||||||||||||||| |||| ||||||||||||||
Db       1195 ACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTGCCACTACAACTCGTACCAAT 1254

Qy       1207 CCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1245
              |||||||||| || || || || || || |||||||||
Db       1255 CCAACTGATTCTATAGATACTGTCGTTGTGCAAGTTCCA 1293
```

```
RESULT 12
CANALS4S1
LOCUS       CANALS4S1               1407 bp    DNA     linear   PLN 16-OCT-1998
DEFINITION  Candida albicans agglutinin-like protein (ALS4) gene, 5' partial
            cds.
ACCESSION   AF024584
VERSION     AF024584.1  GI:3598677
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 1407)
  AUTHORS   Hoyer,L.L., Payne,T.L. and Hecht,J.E.
  TITLE     Identification of Candida albicans ALS2 and ALS4 and localization
            of als proteins to the fungal cell surface
  JOURNAL   J. Bacteriol. 180 (20), 5334-5343 (1998)
  MEDLINE   98440424
   PUBMED   9765564
REFERENCE   2  (bases 1 to 1407)
  AUTHORS   Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1997) Veterinary Pathobiology, University of
            Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES             Location/Qualifiers
     source          1. .1407
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="6"
                     /map="6C"
     gene            1. .>1407
                     /gene="ALS4"
                     /note="agglutinin-like sequence"
                     /allele="ALS4-1"
     CDS             1. .>1407
                     /gene="ALS4"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAC64239.1"
                     /db_xref="GI:3598680"
                     /translation="MLLQFLLLSLCVSVATAKVITGVFNSFNSLTWANAASYPYRGPA
```

```
                     TPTWTAVIGWSLDGATASAGDTFTLDMPCVFKFITDQTSIDLVADGRTYATCNLNSAE
                     EFTTFSSVSCTVTTTMTADTKAIGTVTLPFSFSVGGSGSDVDLANSQCFTAGINTVTF
                     NDGDTSISTTVDFEKSTVASSDRILLSRILPSLSQAVNLFLPQECANGYTSGTMGFST
                     AGTGATIDCSTVHVGISNGLNDWNYPISSESFSYTKTCTSTSVLVTFQNVPAGYRPFV
                     DAYISATRVSSYTMQYTNIYACVGAASVDDSFTHTWRGYSNSQAGSNGITIVVTTRTV
                     TDSTTAVTTLPFNSDTDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPSPNPTVTTTEYWSQSYAT
                     TTTVTAPPGGTDSVIIREP"
     repeat_region   1300. .1407
                     /note="first of multiple copies of 108bp imperfect repeat"
                     /rpt_type=tandem
BASE COUNT      391 a    294 c    253 g    469 t
ORIGIN

  Query Match             58.9%;  Score 733.4;  DB 8;  Length 1407;
  Best Local Similarity   74.5%;  Pred. No. 5.5e-125;
  Matches  923;  Conservative    0;  Mismatches  316;  Indels    0;  Gaps    0;

Qy          7 ATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCT 66
              || ||||||||  |||||||||||| |||||||||| |||||||||||  |||| |
Db         58 ATTACAGGTGTTTTCAATAGTTTTAATTCGTTAACTTGGGCCAATGCTGCTTCTTATCCA 117

Qy         67 TTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACC 126
              |  | ||| ||||     || |||||||  |||||  | || ||||| ||||||||  |
Db        118 TATAGAGGTCCAGCTACTCCTACTTGGACCGCTGTAATAGGATGGTCTTTAGATGGAGCT 177

Qy        127 AGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACT 186
              | ||| | | | || || |||||||| ||  | |||||  ||||| || |||| || ||||
Db        178 ACTGCTAGTGCTGGTGACACATTCACGTTAGACATGCCTTGTGTTTTCAAATTTATTACT 237

Qy        187 TCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTAT 246
                |||||  ||  |||||||   ||| ||||||   ||   ||||||||||| | ||  ||
Db        238 GATCAAACGTCAATTGATTTAGTTGCTGATGGTCGTACTTATGCTACTTGTAATTTGAAT 297

Qy        247 TCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAA 306
              ||||  ||||| || || |||||||||    |  ||||||||||||      |  |||
Db        298 TCTGCCGAAGAGTTTACTACTTTTTCTAGTGTGTCATGTACTGTGACTACTACAATGACT 357

Qy        307 TCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACA 366
              |   ||  || ||  | || || || || |||||  |  |||||| ||| || ||| ||
Db        358 GCTGACACCAAAGCCATAGGAACTGTAACATTACCTTTCTCATTCAGTGTGGGGGGATCA 417

Qy        367 GGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACA 426
              ||||||     |||||||| | ||||| ||||||||||||| || | |||||||||  ||
Db        418 GGTTCAGATGTTGATTTGGCAAATTCTCAATGTTTTACTGCAGGAATCAATACAGTTACT 477

Qy        427 TTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGAT 486
              |||||||||||||| |     || || |     ||||| |||||||| ||||||||  |
Db        478 TTTAATGATGGTGACACTAGCATTTCCACAACAGTTGATTTTGAAAAATCAACCGTGGCC 537

Qy        487 CCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTT 546
              | || |    | | |      || ||| || | || |||||      | |    || ||||
Db        538 TCCAGCGATCGTATCTTGTTGTCAAGAATTTTACCCAGTCTTTCACAAGCAGTAAATCTT 597

Qy        547 TTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGT 606
              ||| |  | | | |||||| |||||||| || |||||||||||||| || || | |  |
Db        598 TTTCTTCCCCAAGAATGTGCAAATGGTTATACTTCTGGTACAATGGGATTTTCGACTGCT 657

Qy        607 AACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAAT 666
                ||  | | |||| ||||| || |   |||||  | || ||  |||| || || |||
Db        658 GGTACTGGTGCTACTATAGATTGTTCCACAGTTCATGTCGGGATATCAAATGGGTTGAAT 717

Qy        667 GATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAAT 726
              |||||||||||||  |||||   ||||| |||  ||||||  || || ||||||||  |
Db        718 GATTGGAATTATCCAATTTCACTGGAATCTTTTTCTTACACAAAGACCTGTACATCAACA 777

Qy        727 GGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTAT 786
              |  ||    | |  |||||||| |||| || ||||||||||  ||||||||||||||||
Db        778 AGTGTTTTAGTAACTTTTCAAAATGTTCCTGCCGGATATCGTCCATTTGTTGATGCTTAT 837

Qy        787 ATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCT 846
              |||||||| |||   || | |  ||||||   |   ||| || |||   ||| |||||| |
Db        838 ATTTCTGCAACACGAGTCAGCTCATATACCATGCAATACACTAATATATATGCTTGTGTT 897

Qy        847 GGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGAT 906
              |||      |         | |  | || |||    |  |||   ||||| |  |||||| |
Db        898 GGCGCGGCTTCTGTTGATGACTCATTTACTCATACTTGGCGGGGATATAGTAATAGTCAA 957

Qy        907 GCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACT 966
              || || ||||| ||||||  |||||| |  ||||||||||||||||||||||||||||||
Db        958 GCTGGTTCTAATGGTATTACCATTGTGGTAACAACTAGAACAGTTACAGACAGTACCACT 1017

Qy        967 GCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTG 1026
              ||||| |||||||||||||||||| |   |  ||| ||||||||||||||||||||||
Db       1018 GCTGTGACTACTTTACCATTCAATTCCGATACTGACAAAACCAAAACAATCGAAATTTTA 1077

Qy       1027 CAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTG 1086
              |||||||||||||| |||||||| |||||||||||||||||||||| |||||||| |||
Db       1078 CAACCTATTCCAACAACTACCATTACAACTTCATATGTTGGTGTGACAACTTCCTACCTG 1137

Qy       1087 ACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACT 1146
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 ACTAAAACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACT 1197

Qy       1147 ACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAAT 1206
              || |||||||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||||
Db       1198 ACTACAACTGTTACCAGTGAATGGACAGGAACAATTACTACCACTACAACTCGTACCAAT 1257

Qy       1207 CCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1245
              |||||||||||  || ||  ||  ||  ||  ||  |||||||||
Db       1258 CCAACTGATTCTATAGATACTGTCGTTGTTCAAGTTCCA 1296


RESULT 13
AF272027
LOCUS       AF272027                4569 bp    mRNA    linear   PLN 24-OCT-2000
DEFINITION  Candida albicans agglutinin-like protein mRNA, partial cds.
ACCESSION   AF272027
VERSION     AF272027.1  GI:10952735
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 4569)
  AUTHORS   Chen,X. and Chen,J.-Y.
  TITLE     ALS4 (agglutinin-like sequence) of Candida albicans
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 4569)
  AUTHORS   Chen,X. and Chen,J.-Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-2000) Shanghai Institute of Biochemistry, Yue
            Yang Road 320, Shanghai 200031, China
FEATURES             Location/Qualifiers
     source          1. .4569
                     /organism="Candida albicans"
                     /mol_type="mRNA"
                     /strain="SC5314"
                     /db_xref="taxon:5476"
     CDS             1. .>4569
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAG25054.1"
                     /db_xref="GI:10952736"
                     /translation="MLLQFLLLSLCVSVATAKVITGIFDSFNSLTWTNAASYSYRGPA
                     NPTWTAVIGWSLDGATASAGDTFTLDMPCVFKFITDQTSIDLVADGRTYATCNLNSAE
                     EFTTFSSVSCTVTTTMTADTKAIGTVTLPFSFSVGGSGSDVDLANSQCFTAGINTVTF
                     NDGDTSISATVDFEKSTVASSDRILLSRILPSLSQAVSLFLPQECANGYTSGTMGFST
                     AGTGATIDCSTVHVGISNGLNDWNYPISSESFSYTKTCTSTSVLVTYQNVPAGYRPFV
                     DAYVSATRVSSYAMRYTNIYACVGAASVDDSFTHTWSGYSNSQAGSNGITIVVTTRTV
                     TDSTTAVTTLPFNSESDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPSPNPTVTTTEYWSQSYAT
                     TTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNP
                     TVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPP
                     GGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYW
```

SQSFATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVII
REPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYVTTS
TITAPPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTV
TTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNYTVTTTEYWSQSYATTTTVTAPPGG
TDTVIIREPPSPTVTTTEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQ
SYATTTTVTSPPGGTDIVIIREPPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIRE
PPNPTVTTTEYWSQSYATTTTVTAPPGGTDSVIIREPPNPTVTTTEYWSQSYATTTTV
TAPPGGTDSVIIREPPNPTVTTTEYWSQSFATTTTVTAPPGGTDSVIIREPPNPTVTT
TEYWSQSYATTTTVTAPPGGTDTVIIREPPNPTVTTTEYWSQSYATTLTITAPPGGTN
SVIIRVHSSTNDESSESTPSTLSVPSFSGSISIVSTVSRPHYVNSTVTHLPSSSSKPV
DIPSSDVVPSTNDNSLTSLTGSENGETSVAISTTFCDDENGCQTSIPQGSVVRTTAPT
TATPTPIIGDNNGSGKSKSGELSSTGSVTTNTATPDVPSTKVPSNPGAPGTGVPPPLA
PSTETQTTNNVPGSPNIPATGTTDIIRESTTVSHTVTGNGNTGVPMNPNPALTTSTSL
TGATNSATNPSHETGVNTGSRGSTNIVTPPSSATATVVIPGTDNGATTKGQDTAGGNS
NGFTATTNTQGGNNEPGNQPGTNTTGEPVGTTGTQSVESTSQPTTLSQQTTSSLISTP
LASTFDGSGSIVQHSGWLYVLLTAISIFF"

BASE COUNT 1318 a 1186 c 787 g 1278 t
ORIGIN

Query Match 58.9%; Score 733; DB 8; Length 4569;
Best Local Similarity 74.3%; Pred. No. 4.9e-125;
Matches 925; Conservative 0; Mismatches 320; Indels 0; Gaps 0;

```
Qy          1 AAGACAATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAAT 60
Db         52 AAGGTTATTACGGGCATTTTCGATAGTTTTAATTCGTTAACTTGGACCAATGCTGCTTCT 111

Qy         61 TATGCTTTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGAT 120
Db        112 TATTCATATAGAGGTCCAGCTAATCCTACTTGGACCGCTGTAATAGGATGGTCTTTAGAT 171

Qy        121 GGTACCAGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATAT 180
Db        172 GGAGCTACTGCTAGTGCTGGTGACACATTCACGTTAGACATGCCATGTGTTTTCAAATTT 231

Qy        181 ACTACTTCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAA 240
Db        232 ATTACTGATCAAACGTCAATTGATTTAGTTGCTGATGGTCGTACTTATGCTACTTGTAAT 291

Qy        241 TTTTATTCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCT 300
Db        292 TTGAACTCTGCCGAAGAGTTTACTACTTTTTCTAGTGTGTCATGTACTGTGACTACTACA 351

Qy        301 TTGAAATCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGT 360
Db        352 ATGACTGCTGACACCAAAGCCATAGGAACTGTAACATTACCTTTCTCATTCAGTGTGGGG 411

Qy        361 GGAACAGGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACA 420
Db        412 GGATCAGGTTCAGATGTTGATTTGGCAAATTCTCAATGTTTTACTGCAGGAATCAATACA 471

Qy        421 GTCACATTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACC 480
Db        472 GTTACTTTTAATGATGGTGACACTAGCATTTCCGCAACAGTTGATTTTGAAAAATCAACC 531

Qy        481 GTTGATCCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACA 540
Db        532 GTGGCCTCCAGCGATCGTATCTTGTTGTCAAGAATTTTACCCAGTCTTTCACAAGCAGTA 591

Qy        541 ACTCTTTTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCC 600
Db        592 AGTCTTTTTCTTCCCCAAGAATGTGCAAATGGTTATACTTCTGGTACAATGGGATTTTCA 651

Qy        601 AGTAGTAACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGA 660
Db        652 ACTGCTGGTACTGGTGCTACTATAGATTGTTCCACAGTTCATGTCGGGATATCAAATGGG 711

Qy        661 TTAAATGATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACA 720
Db        712 TTGAATGATTGGAATTATCCAATTTCACTGGAATCTTTTTCTTACACAAAGACCTGTACA 771

Qy        721 TCTAATGGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGAT 780
```

```
Db        772 TCAACAAGTGTTTTAGTAACATATCAGAATGTTCCTGCCGGATATCGTCCATTTGTCGAT 831

Qy        781 GCTTATATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACT 840
Db        832 GCTTATGTTTCTGCAACACGAGTCAGCTCATATGCTATGAGATACACTAATATATATGCT 891

Qy        841 TGTGCTGGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAAT 900
Db        892 TGTGTTGGTGCGGCTTCTGTTGATGACTCATTTACTCATACTTGGCTGGGATATAGTAAT 951

Qy        901 AGTGATGCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGT 960
Db        952 AGTCAAGCTGGTTCTAATGGTATTACCATTGTGGTAACAACTAGAACAGTTACAGACAGT 1011

Qy        961 ACCACTGCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAA 1020
Db       1012 ACCACTGCTGTGACTACTTTACCATTCAATTCCGAATCTGACAAAACCAAAACAATCGAA 1071

Qy       1021 ATTTTGCAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCC 1080
Db       1072 ATTTTACAACCTATTCCAACAACTACCATTACAACTTCATATGTTGGTGTGACAACTTCC 1131

Qy       1081 TATCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1140
Db       1132 TACCTGACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATAT 1191

Qy       1141 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGT 1200
Db       1192 CATACTACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACTACAACTCGT 1251

Qy       1201 ACCAATCCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1245
Db       1252 ACCAATCCAACTGATTCTATAGATACTGTCGTTGTGCAAGTTCCA 1296
```

```
RESULT 14
AF202529
LOCUS       AF202529                1007 bp    DNA     linear   PLN 13-DEC-2000
DEFINITION  Candida dubliniensis strain CD36 agglutinin-like protein Alsd2p
            (ALSD2) gene, partial cds.
ACCESSION   AF202529
VERSION     AF202529.1  GI:11493723
KEYWORDS    .
SOURCE      Candida dubliniensis
  ORGANISM  Candida dubliniensis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 1007)
  AUTHORS   Hoyer,L.L., Hecht,J.E., En,J., Kapteyn,J.C. and Klis,F.M.
  TITLE     Evidence suggesting the presence of an ALS gene family in Candida
            dubliniensis and Candida tropicalis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1007)
  AUTHORS   Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-1999) Veterinary Pathobiology, University of
            Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES             Location/Qualifiers
     source          1. .1007
                     /organism="Candida dubliniensis"
                     /mol_type="genomic DNA"
                     /strain="CD36"
                     /db_xref="taxon:42374"
     gene            <1. .>1007
                     /gene="ALSD2"
     mRNA            <1. .>1007
                     /gene="ALSD2"
                     /product="agglutinin-like protein Alsd2p"
     CDS             <1. .>1007
                     /gene="ALSD2"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein Alsd2p"
```

```
                     /protein_id="AAG35623.2"
                     /db_xref="GI:11693480"
                     /translation="ASPGDTFTLIMPCVFKFITTQTSVDLTANGVKYATCSFYSGEEF
                     TTFSSLTCTVNSALTSSVKAFGTVTLPISFNVGGTGSSVDLEDSKCFTAGTNTVTFTD
                     GDNKVSTTVDFEKSTVDSTGYLTSSRLMPSLNKVTSLFVAPQCARGYTSGTIGFSSSN
                     GGVSFDCSNVHVGITNGVNDWNFPVSSESFSYTKTCSSTGITITYRNVPAGYRPFIDA
                     YITASDVNSYTLSYTNDYTCVGGSVQHKPFTLRWSGYKNGEAGSNGIVIVATTRTVTD
                     STTAVTTLPFNSDVDKTKTIEILQPIPTTTITTLYVGVTTSYSTKTAPIGDTATLFID
                     IP"
BASE COUNT      292 a    205 c    176 g    334 t
ORIGIN

  Query Match            58.8%;  Score 731.8;  DB 8;  Length 1007;
  Best Local Similarity  82.9%;  Pred. No. 1.2e-124;
  Matches  835;  Conservative    0;  Mismatches  172;  Indels    0;  Gaps    0;

Qy        130 GCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACTTCA 189
              || | ||| || || ||||| ||||||| |||||  |||||||| ||||  |  ||| |
Db          1 GCAAGTCCGGGAGACACATTTACATTGATCATGCCTTGTGTGTTCAAATTCATAACTACC 60

Qy        190 CAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTATTCT 249
              |||||||||||||||||| || ||| |||||||||||||||||||| ||||  ||||||||
Db         61 CAAACATCTGTTGATTTGACCGCCAATGGTGTTAAATATGCTACATGTCTGTTTTATTCG 120

Qy        250 GGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAATCA 309
              ||||||||||||||||||||  |||||||||||||||||| |||   ||||| |  |||
Db        121 GGTGAAGAATTCACAACTTTTTCATCATTAACATGTACTGTTAACAGTGCTTTAACCTCA 180

Qy        310 TCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACAGGT 369
              ||  ||||||| |||||||| |||||||||| ||  |||||||||| |||||||| ||
Db        181 TCTGTTAAGGCTTTTGGTACCGTTACTTTACCCATCTCATTCAATGTAGGTGGAACGGGA 240

Qy        370 TCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACATTT 429
              || |||  |||||| ||||| || ||||| || |||||||||||||| || ||||| |||
Db        241 TCGTCAGTTGATTTAGAAGACTCCAAATGCTTCACTGCTGGTACCAACACCGTCACTTTT 300

Qy        430 AATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGATCCA 489
              |  || || || ||  |  |||||| |  |||||| |||||||| ||||| |||||| ||
Db        301 ACCGACGGCGACAACAAAGTCTCAACTACTGTTGATTTTGAAAAATCAACAGTTGATTCA 360

Qy        490 AGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTTTTT 549
              | || |||||||  | | |||||| |||||||||||||||||||| |||||| ||||||||
Db        361 ACTGGATATTTGACTTCCTCCAGACTTATGCCAAGTCTCAATAAAGTCACATCTCTTTTT 420

Qy        550 GTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGTAAC 609
              || |||||||||||| ||  ||||| |||||||||||||| || ||  || ||||||||
Db        421 GTAGCACCACAATGTGCAAGAGGTTATACATCTGGTACAATTGGATTTTCGAGTAGTAAT 480

Qy        610 GGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAATGAT 669
              ||||  ||| || |||| || |||||| ||||| |||| || ||||| ||  |  ||||||
Db        481 GGTGGTGTTTCTTTTGACTGTTCAAATGTTCATGTTGGAATTACAAATGGTGTGAATGAT 540

Qy        670 TGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAATGGA 729
              ||||||| ||| || ||||| |||||||||||||||||||||||| | ||||  ||
Db        541 TGGAATTTTCCAGTGTCATCAGAATCATTTAGTTACACTAAAACTTGTTCGTCTACAGGT 600

Qy        730 ATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTATATT 789
              |||   || | |||   |||||| |||||||||||||||||||||| || ||||| |||
Db        601 ATTACAATCACATACAGAAATGTTCCTGCTGGTTATCGTCCATTTATCGACGCTTACATT 660

Qy        790 TCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCTGGC 849
              |||||  | ||||||||||  |||||  ||  |||||||||||||||||||||||| |||
Db        661 ACTGCTTCGGATGTTAACCTGTATACCTTGTCATATACCAATGATTATACTTGTGTTGGT 720

Qy        850 AGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGATGCC 909
              || |  |||||  |||||| || |||||||||||  |||||||||| ||| | || |||
Db        721 GGTAGCGTGCAACATAAACCATTTACTTTAAGATGGTCTGGATACAAAAATGGGGAAGCC 780

Qy        910 GGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACTGCT 969
              || |||||||||||||||||||||||||||||| |||||||||||||| |||||  ||||||
Db        781 GGGTCTAACGGTATTGTCATTGTTGCTACAACAAGAACAGTTACAGATAGTACTACTGCT 840

Qy        970 GTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTGCAA 1029
              || |||||||||||||||||  |||||| |||||  |||||||| ||||||||||||||
Db        841 GTAACTACTTTACCATTCAATTCCGATGTTGACAAAACTAAAACAATTGAAATTTTGCAA 900

Qy       1030 CCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTGACT 1089
              |||||||||||  |||||||||||||| | ||| |||||||||||||||||| || |||||
Db        901 CCTATTCCAACAACTACCATCACAACATTATACGTTGGTGTGACTACTTCTTACCTGACC 960

Qy       1090 AAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCC 1136
              || |||||||||||||||| ||||||||| || || ||||| | ||
Db        961 AAAACTGCACCAATTGGTGACACAGCTACTCTTTTTATTGATATTCC 1007


RESULT 15
CANALS42S1
LOCUS       CANALS42S1              1407 bp    DNA     linear   PLN 16-OCT-1998
DEFINITION  Candida albicans agglutinin-like protein (ALS4) gene, 5' partial
            cds.
ACCESSION   AF024586
VERSION     AF024586.1  GI:3598682
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 1407)
  AUTHORS   Hoyer,L.L., Payne,T.L. and Hecht,J.E.
  TITLE     Identification of Candida albicans ALS2 and ALS4 and localization
            of als proteins to the fungal cell surface
  JOURNAL   J. Bacteriol. 180 (20), 5334-5343 (1998)
  MEDLINE   98440424
   PUBMED   9765564
REFERENCE   2  (bases 1 to 1407)
  AUTHORS   Hoyer,L.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1997) Veterinary Pathobiology, University of
            Illinois, 2001 S. Lincoln Avenue, Urbana, IL 61802, USA
FEATURES             Location/Qualifiers
     source          1. .1407
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /strain="1161"
                     /db_xref="taxon:5476"
                     /chromosome="6"
                     /map="6C"
     gene            1. .>1407
                     /gene="ALS4"
                     /note="agglutinin-like sequence"
                     /allele="ALS4-2"
     CDS             1. .>1407
                     /gene="ALS4"
                     /codon_start=1
                     /transl_table=12
                     /product="agglutinin-like protein"
                     /protein_id="AAC64241.1"
                     /db_xref="GI:3598685"
                     /translation="MLLQFLLLSLCVSVATAKVITGVFNSFNSLTWANAASYPYRGPA
                     TPTWTAVIGWSLDGATASAGDTFTLDMPCVFKFITDQTSIDLVADGRTYATCNLNSAE
                     EFTTFSSVSCTVTTTMTADTKAMGTVTLPFSFSVGGSGSDVDLANSQCFTAGINTVTF
                     NDGDTSISTTVDFEKSTVASSDRILLSRILPSLSQAVSLFLPQECANGYTSGTMGFST
                     AGTGATIDCSTVHVGISNGLNDWNYPISSESFSYTKTCTSTSVLVTYQNVPAGYRPFV
                     DAYVSATRVSSYTMQYTNIYACVGAASVDDSFTHTWRGYSNSQAGSNGITIVVTTRTV
                     TDSTTAVTTLPFNSDTDKTKTIEILQPIPTTTITTSYVGVTTSYSTKTAPIGETATVI
                     VDVPYHTTTTVTSEWTGTITTTTTRTNPTDSIDTVVVQVPSPNPTVTTTEYWSQSYAT
                     TTTVTAPPGGTDSVIIREP"
     repeat_region   1300. .1407
                     /note="first of multiple copies of 108bp imperfect repeat"
                     /rpt_type=tandem
BASE COUNT      389 a    297 c    257 g    464 t
ORIGIN

  Query Match            58.5%;  Score 728.6;  DB 8;  Length 1407;
  Best Local Similarity  74.3%;  Pred. No. 4.2e-124;
```

```
  Matches  920;  Conservative    0;  Mismatches  319;  Indels    0;  Gaps    0;

Qy          7 ATCACTGGTGTTTTTGATAGTTTTAATTCATTAACTTGGTCCAATGCTGCTAATTATGCT 66
              ||  || ||||||||  ||||||||||||| |||||||| ||||||||||  |||| |
Db         58 ATTACAGGTGTTTTCAATAGTTTTAATTCGTTAACTTGGGCCAATGCTGCTTCTTATCCA 117

Qy         67 TTCAAAGGGCCAGGATACCCAACTTGGAATGCTGTTTTGGGTTGGTCCTTAGATGGTACC 126
              |  | ||| ||||     || |||||||  |||||  | || ||||| ||||||||  |
Db        118 TATAGAGGTCCAGCTACTCCTACTTGGACCGCTGTAATAGGATGGTCTTTAGATGGAGCT 177

Qy        127 AGTGCCAATCCAGGGGATACATTCACATTGAATATGCCATGTGTGTTTAAATATACTACT 186
              | ||| | | | || || |||||||| ||  | ||||| ||||| || |||| || ||||
Db        178 ACTGCTAGTGCTGGTGACACATTCACGTTAGACATGCCTTGTGTTTTCAAATTTATTACT 237

Qy        187 TCACAAACATCTGTTGATTTAACTGCCGATGGTGTTAAATATGCTACTTGTCAATTTTAT 246
                ||||| ||  |||||||  ||| ||||||  ||  |||||||||||| | ||  |
Db        238 GATCAAACGTCAATTGATTTAGTTGCTGATGGTCGTACTTATGCTACTTGTAATTTGAAC 297

Qy        247 TCTGGTGAAGAATTCACAACTTTTTCTACATTAACATGTACTGTGAACGACGCTTTGAAA 306
              ||||  ||||| || || ||||||||||   |  ||||||||||||      |  |||
Db        298 TCTGCCGAAGAGTTTACTACTTTTTCTAGTGTGTCATGTACTGTGACTACTACAATGACT 357

Qy        307 TCATCCATTAAGGCATTTGGTACAGTTACTTTACCAATTGCATTCAATGTTGGTGGAACA 366
              |   ||  || ||  | || || || || |||||  |  |||||| ||| || ||| ||
Db        358 GCTGACACCAAAGCCATGGGAACTGTAACATTACCTTTCTCATTCAGTGTGGGGGGATCA 417

Qy        367 GGTTCATCAACTGATTTGGAAGATTCTAAATGTTTTACTGCTGGTACCAATACAGTCACA 426
              ||||||     |||||||| | ||||| ||||||||||||| || | ||||||||| ||
Db        418 GGTTCAGATGTTGATTTGGCAAATTCTCAATGTTTTACTGCAGGAATCAATACAGTTACT 477

Qy        427 TTTAATGATGGTGATAAAGATATCTCAATTGATGTTGAGTTTGAAAAGTCAACCGTTGAT 486
              |||||||||||||| |      || || |     ||||| |||||||| ||||||||
Db        478 TTTAATGATGGTGACACTAGCATTTCCACAACAGTTGATTTTGAAAAATCAACCGTGGCC 537

Qy        487 CCAAGTGCATATTTGTATGCTTCCAGAGTTATGCCAAGTCTCAATAAGGTCACAACTCTT 546
              | || |    | | |      || ||| || | || |||||     | |    || ||||
Db        538 TCCAGCGATCGTATCTTGTTGTCAAGAATTTTACCCAGTCTTTCACAAGCAGTAAGTCTT 597

Qy        547 TTTGTGGCACCACAATGTGAAAATGGTTACACATCTGGTACAATGGGGTTCTCCAGTAGT 606
              ||| |  | | | |||||| ||||||||| || |||||||||||||| || || | |  |
Db        598 TTTCTTCCCCAAGAATGTGCAAATGGTTATACTTCTGGTACAATGGGATTTTCAACTGCT 657

Qy        607 AACGGTGACGTTGCTATTGATTGCTCAAATATTCATATTGGTATCACAAAAGGATTAAAT 666
                ||  | | |||| ||||| || |   ||||| | || ||  |||| || || |||
Db        658 GGTACTGGTGCTACTATAGATTGTTCCACAGTTCATGTCGGGATATCAAATGGGTTGAAT 717

Qy        667 GATTGGAATTATCCGGTTTCATCTGAATCATTTAGTTACACTAAAACTTGTACATCTAAT 726
              |||||||||||||  |||||   ||||| |||  |||||| || || |||||||| |
Db        718 GATTGGAATTATCCAATTTCACTGGAATCTTTTTCTTACACAAAGACCTGTACATCAACA 777

Qy        727 GGAATTCAGATTAAATATCAAAATGTACCTGCTGGTTATCGTCCATTTATTGATGCTTAT 786
              | ||    | |  ||||| ||||| ||||| || |||||||||||| | |||||||||
Db        778 AGTGTTTTAGTAACTTATCAGAATGTTCCTGCCGGATATCGTCCATTTGTCGATGCTTAT 837

Qy        787 ATTTCTGCTACAGATGTTAACCAATATACTTTAGCATATACCAATGATTATACTTGTGCT 846
              |||||||  |||   || | |  ||||||   |   ||| || |||   ||| ||||||
Db        838 GTTTCTGCAACACGAGTCAGCTCATATACCATGCAATACACTAATATATATGCTTGTGTT 897

Qy        847 GGCAGTCGTCTGCAAAGTAAACCTTTCACTTTAAGATGGACTGGATACAAGAATAGTGAT 906
              |||      |         | |  | || |||    |  |||   ||||| |  |||||
Db        898 GGCGCGGCTTCTGTTGATGACTCATTTACTCATACTTGGCGGGGATATAGTAATAGTCAA 957

Qy        907 GCCGGATCTAACGGTATTGTCATTGTTGCTACAACTAGAACAGTTACAGACAGTACCACT 966
              || || ||||| ||||||  |||||| |  ||||||||||||||||||||||||||||
Db        958 GCTGGTTCTAATGGTATTACCATTGTGGTAACAACTAGAACAGTTACAGACAGTACCACT 1017

Qy        967 GCTGTCACTACTTTACCATTCAATCCAAGTGTTGATAAAACCAAAACAATCGAAATTTTG 1026
              ||||| |||||||||||||||| |    |  ||| ||||||||||||||||||||||
Db       1018 GCTGTGACTACTTTACCATTCAATTCCGATACTGACAAAACCAAAACAATCGAAATTTTA 1077

Qy       1027 CAACCTATTCCAACCACTACCATCACAACTTCATATGTTGGTGTGACTACTTCCTATCTG 1086
              |||||||||||||| |||||||| |||||||||||||||||||| |||||||| |||
Db       1078 CAACCTATTCCAACAACTACCATTACAACTTCATATGTTGGTGTGACAACTTCCTACCTG 1137

Qy       1087 ACTAAGACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACT 1146
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1138 ACTAAAACTGCACCAATTGGTGAAACAGCTACTGTTATTGTTGATGTGCCATATCATACT 1197

Qy       1147 ACCACAACTGTTACCAGTGAATGGACAGGAACAATCACTACCACCACAACTCGTACCAAT 1206
              || |||||||||||||||||||||||||||||||| |||||||| ||||||||||||||
Db       1198 ACTACAACTGTTACCAGTGAATGGACAGGAACAATTACTACCACTACAACTCGTACCAAT 1257

Qy       1207 CCAACTGATTCAATTGACACAGTGGTGGTACAAGTTCCA 1245
              |||||||||| || || || || || || |||||||||
Db       1258 CCAACTGATTCTATAGATACTGTCGTTGTTCAAGTTCCA 1296

Search completed: January 17, 2004, 22:42:32
Job time : 4790 secs
```